



## Research article

# Evaluating human resources management literacy: A performance analysis of ChatGPT and bard

Raghu Raman<sup>\*</sup>, Murale Venugopalan, Anju Kamal

Amrita School of Business, Amrita Vishwa Vidyapeetham, Amritapuri, India

## ARTICLE INFO

## Keywords:

Human resource management  
LLM  
Generative AI  
Text mining  
HR policy  
Hiring  
Ethics  
Managerial decisions

## ABSTRACT

This study presents a comprehensive analysis comparing the literacy levels of two Generative Artificial Intelligence (GAI) tools, ChatGPT and Bard, using a dataset of 134 questions from the Human Resources (HR) domain. The generated responses are evaluated for accuracy, relevance, and clarity. We find that ChatGPT outperforms Bard in overall accuracy (84.3% vs. 82.8%). This difference in performance suggests that ChatGPT could serve as a robotic advisor in transactional HR roles. In contrast, Bard may possess additional safeguards against misuse in the HR function, making it less capable of generating responses to certain types of questions. Statistical tests reveal that although the two systems differ in their mean accuracy, relevance, and clarity of the responses, the observed differences are not always statistically significant, implying that both tools may be more complementary than competitive. The Pearson correlation coefficients further support this by showing weak to non-existent relationships in performance metrics between the two tools. Confirmation queries don't improve ChatGPT or Bard's response accuracy. The study thus contributes to emerging research on the utility of GAI tools in Human Resources Management and suggests that involving certified HR professionals in the design phase could enhance underlying language model performance.

## 1. Introduction

Since the second half of the last decade, Artificial Intelligence (AI) tools, platforms, and those claiming 'intelligent status have become vital elements of business entities and the society we belong to. According to Ref. [1], AI can automate organizational procedures, extract valuable insights from large datasets, provide predictive analytics, and transcend human analytical skills. AI plays a pivotal role in assisting firms in minimizing the time and expenses related to the decision-making process, enhancing operational effectiveness, and delivering superior quality customer experience. The OpenAI [2] discusses the design and development of generative AI models, including ChatGPT and its subsequent version, ChatGPT-4, as a noteworthy example of the ongoing progress in AI.

Major technological corporations like Microsoft, Google, Meta, and others are making substantial financial investments and perpetrating valuable resources to the Generative AI (GAI) tools industry, augmenting its rapid expansion. These technological developments offer the capacity to improve the effectiveness of GAI-based applications. However, they also raise apprehensions regarding their capability to generate convincing responses that ultimately deceive the user. This is supplemented by the fact that a GAI model's ability to offer real-time information is constrained by its dependence on past training data, specifically, GPT-4's dataset, which is only updated until 2021. Furthermore, the lack of transparency inherent in these algorithms and their capacity to propagate

<sup>\*</sup> Corresponding author. Amrita School of Business, Amritapuri, India.

E-mail addresses: [raghu@amrita.edu](mailto:raghu@amrita.edu) (R. Raman), [v\\_murale@asb.kochi.amrita.edu](mailto:v_murale@asb.kochi.amrita.edu) (M. Venugopalan), [anjukamal@am.amrita.edu](mailto:anjukamal@am.amrita.edu) (A. Kamal).

erroneously presents significant concerns regarding reliability and ethical implications. Google encountered a significant financial setback in 2023, which can be linked to the deceptive response provided by its GAI chatbot, Bard.

In the past few years, there has been a discernible rise in the prevalence of research in the Human resource management domain with a specific focus on GAI [3–5]. Multiple studies have shown evidence that AI holds promise in improving diversity and streamlining the recruitment process [6]), enhancing our understanding of AI adoption trends among Human Resources professionals and promoting greater openness in AI algorithms [4,5,7]. In the past, Strategic Human Resource Management has emphasized ensuring that an organization's human resources are effectively aligned with its long-term objectives [8]). There is a growing discussion among academics worldwide about the potential effects of GAI and related technologies on contemporary human resource management practices.

Rhem [9] asserts that Machine learning techniques are important in AI and business because they can accurately represent intricate systems by discerning patterns within extensive datasets. The study by Ref. [10] examines the research state of artificial intelligence (AI) and machine learning (ML) applications in the Banking, Financial Services, and Insurance sectors.

The promising prospects of AI generate a sense of optimism owing to its remarkable capacity for adaptability and efficiency. The current learning methodology necessitates revision. AI datasets and algorithms are prone to human biases, leading to inaccurate results. Similar persistent concerns exist regarding the possibility of divesting ethical standards in managerial decision-making and the gradual decline of professional skills. Like its predecessors, AI exhibits certain limitations. The challenges mentioned above result from the methodologies and datasets utilized in training AI models [11].

As the advancement of AI continues, it becomes increasingly apparent that AI models will be increasingly deployed, especially in human resources and recruitment [12,13]. The possibility of bias in Ref. [14] and especially within the realm of GAI is a matter of tremendous concern [15]. The Amazon AI recruiter serves as a concrete example of how bias becomes evident in real-world scenarios; based on prior research conducted by Kaplan and Haenlein [16] as well as Taniguchi et al. [17], it has been observed that the AI system demonstrated a tendency to favor male candidates. This exemplifies the importance of conducting testing and implementing modifications throughout the various phases of implementing GAI systems. The importance of bias and ethics has always been the highest priority in the academic field focused on studying human subjects. Organizations on a global scale must comprehensively understand the benefits and drawbacks of using GAI-based solutions within HRM [18].

The Human Resources function has traditionally been driven by employer instructions to achieve organizational goals. Nevertheless, the new world of work tends to adopt methods that allow employees to access Human Resources anytime, anywhere, and where human capital is treated as a shared responsibility between employers and employees [19]). The demand for individual Human Resource Development is growing in such a workplace, where employees can access Human Resource Development possibilities according to their interests. GAI tools like ChatGPT and Bard can provide employees with human capital development resources like online tutorials and training modules and provide personalized feedback for skill development. Hence, the purpose of our study is to analyze ChatGPT and Bard's HR literacy performance to gain insight into their strengths and weaknesses and what HR roles they would be most suitable for.

Incorporating GAI, such as ChatGPT and Bard, has become a more prevalent area of investigation in recent years. However, not many studies have been conducted in the realm of Human Resource Management (HRM), which warrants the need for additional empirical research to investigate their effectiveness and proficiency in people management operations, especially concerning Human Resource Management literacy [5]. The application of GAI in HRM operations has the potential to amplify effectiveness and can extend support in managerial decisions while evaluating employee performance. A review of extant studies highlights impending challenges associated with human-focused activities while utilizing GAI due to ethical dimensions related to data protection, permission, potential litigation, and biases that are innate to these artificial intelligence-based solutions [4,20]. Although GAI-based tools exhibited remarkable accomplishments in various domains, especially in automating scientific methods to a more significant extent, a certain degree of trepidation remains concerning their pertinency in HR-related aspects. These apprehensions are mainly because of the reason that HR entails comprehending individuals and organizational structures. The objective of this research is to address the prevailing knowledge gap by carrying out a comparative analysis of the HR literacy levels amongst between two conspicuous GAI tools, namely ChatGPT and Bard. This analysis is based on a dataset obtained from queries answered by GAI models for Strategic Human Resources Management (HRM) Professional Certification programs, which have yet to receive adequate attention within the field of Human Resource Management. We compare the performance of GAI tools based on the accuracy, relevance, and clarity of the generated responses and use statistical tests to assess the response differences. We deploy the cosine similarity metric [21] to understand the similarities of the responses and the readability of responses to measure the differences in the difficulty of understanding the responses between the GAI tools. In addition, we probe into the potential biases, constraints, and broader corollaries these tools might possess. The findings of our study will help in garnering evidence about the efficacy of these GAI tools in the field of HRM, especially regarding ethical aspects, cultural competence, and alignment with organizational requirements and expectations of stakeholders.

## 2. Literature review

### 2.1. GAI tools in human resources management

GAI tools that provide human-like conversations shift control and knowledge from people to machines, affecting human beings as well as organizations [22,23]. Human Resource Management studies investigate Generative AI [4] using a socio-technical approach and a stakeholder approach [24]. The socio-technical approach emphasizes how work design affects stakeholder outcomes. The stakeholder approach considers various stakeholders, such as shareholders, management, and employees, for organizational

effectiveness.

Studies adopting a socio-technical approach examine how HRM functions improve routine processes with GAI. Edlich et al. [25] explored how GAI chatbots are used in HR services to administer benefits, answer employee inquiries, and maintain records. Bhatt & Muduli [26] and Tcharnatsky & Vogt [27] investigate GAI in training programs and assert that technology-aided learning through natural language processing and interactive voice response can improve learning efficiency. GAI drives employment relations [28] and web-based recruiting systems [29,30]. In the performance management function, GAI maps the performance of the individuals with the objectives of the organization [31], processes complex data in less time [3,32], and writes reviews [33]. GAI identifies challenges and suggests solutions to promote sustainable HRM practices [5,34]. The jobs augmented by GAI would require employees to train AI systems, evaluate outputs and manage policies of AI systems [35,36].

Studies adopting a stakeholder approach contend that GAI may positively or negatively influence employees. Chowdhury et al. [5] investigated the turnover of employees and, through the lens of the resource-based theory of firms, emphasized the need for transparent algorithms in GAI, as argued by Raisch and Krakowski [37]. Though Fitzpatrick et al. [38] deployed a GAI chatbot to provide mental health support for employee well-being and concluded that the symptoms of depression significantly reduced in the participants of the treatment group, there exists a possibility that the integration of GAI into routine work could increase anxiety about job security [2] and affect employee wellbeing. The LLMs of GAI can augment human efforts through various ways like writing and upskilling [2,39]. Since perceived satisfaction from efforts is related to engagement [39], adopting GAI can enhance employee engagement. However, redefining roles, new processes, and job displacement could also affect employee engagement in OpenAI [2].

Taken together, these studies suggest that using GAI as a decision-support tool in HRM improves the efficiency of HR processes. However, such an approach would lack the ability to comprehend nuances of specific circumstances like GAI's focus on economic and societal impact [30], quality of human resources, team management, and diverse expectations of team members [40]. When used in HRM, GAI presents some concerns, like its consequences for science and society [23,41], the impact on practices of law [42], and collective action that subsequently impacts management theories [43]. The fact that GAI is trained mainly on large datasets in English could make other languages less important and obliterate diverse thinking since language is contextualized in a culture [44]. Misusing GAI raises concerns about data usage privacy and consent [4] and legal problems [45]. GAI is prone to biases [46] that could be attributed to the LLM training data [47,48]. Ethical usage should be emphasized [49] Since GAI prompts human beings to engage in unethical behaviors [50].

Studies have compared the performance of GAI tools in different settings.

### 2.1.1. Applied sciences

Rao et al. [51] proposed a qualitative framework for the Myers–Briggs Type Indicator [52] by comparing the performance of LLMs in evaluating human personalities. Patnaik & Hoffmann [53], who investigated the hallucinations of LLMs in clinical medicine - on patients' view of anesthesia before surgery, found that ChatGPT exhibited intellectually superior performance over Bard in psychiatry. Patil et al. [54] compare the radiology knowledge of ChatGPT and Bard and conclude that both display reasonable radiology knowledge and should be used with conscious knowledge of their limitations. Studies have compared the performance of the LLMs to generate accurate responses to common myopia-related queries [55] to differentiate between medical emergency and non-emergency [56], to generate differential diagnoses in neurodegenerative disorders [57], and to generate recommendations that reduce meningitis outbreaks [58].

### 2.1.2. Automation

In problem-solving through algorithms, Alexander [59] compared ChatGPT and Bard in competitive tasks for their accuracy of codes, test code, and specifications. The author concludes that while ChatGPT provided accurate responses, Bard leveraged Python libraries and generated quick solutions. Though ChatGPT-aided robotics like RobotGPT [60] have been implemented in several cases [61,62], Bard-aided robotics is in its infancy [63]. In the MITRE framework of cyber security that organizations use to understand security readiness and existing vulnerabilities, ChatGPT outperforms Bard in time series analysis and in relevant code generation that aligns with MITRE [64] Ahmed et al. [63] compared Bard and ChatGPT in use cases for translation, product specifications, and transcripts. They assert that ChatGPT identifies accurate improvement areas in electronics while in customer service. It saves time and enhances experiences through personalized interactions. While Hans [65] compares ChatGPT and Bard in code generation and concludes that both exhibit similar levels of consistency in their performance, Tafferfer et al. [66] conclude that Bard generates creative language that finds applications in marketing, publishing, writing, and advertising and performs superior in code development and error detection in programming languages compared to ChatGPT. In the context of automating the journal quality evaluation, Dadkhah et al. [67] contend that ChatGPT is an unreliable tool for detecting hijacked and predatory journals and suggests an alternative tool. McGowan et al. [68] compared ChatGPT and Bard in English, Spanish, and Italian to produce accurate references that supplement academic literature search and contend that the legitimacy and accuracy are questionable and the outputs of both LLMs can be verified independently.

### 2.1.3. Neutrality

West [69] observed that while Bard condemned the invasion of Ukraine by Russia, ChatGPT took a neutral stand. Bard did not respond to topics like the Muslim minorities in China or the holocaust, but ChatGPT provided neutral responses. In the context of the TikTok ban, while ChatGPT provided historical reasoning and Donald Trump's attempts to ban it, Bard focused on the population base of youth who used TikTok and the impact of the ban on the US economy. When prompted with political questions about Trump and Biden, Bard generated more opinionated responses than ChatGPT [69].

#### 2.1.4. Learning

Plevris et al. [70] assert that ChatGPT-4 wins over ChatGPT3.5 on correctness, while Bard outperforms ChatGPT-4 in answering questions published online. Bard creates customized learning materials and uses Google Workspace and Google Classroom, which ChatGPT cannot access [71,72]. Dao [73] compared the performance of ChatGPT, Microsoft Bing Chat, and Google Bard on English language proficiency at high schools in Vietnam and concluded that the LLMs outperform the students and Bing is the best among the three. Santos [74] compared the performance of ChatGPT-3.5, ChatGPT-4, Bing Chat, and Bard on Physics problems. Despite the disparities, he asserts they can all think critically, solve problems, comprehend concepts, personalize learning, and display subject knowledge. Raman et al. [75] offer valuable insights into ChatGPT on policy, information sciences, education, scientific publishing, and biomedical sciences by examining the early attention of ChatGPT research using the Altmetric Attention Score. An accelerating role for AI tools in expediting academic publication is further established by Raman [76] by systematically analyzing the acknowledgment of ChatGPT in publications. The study contends that acknowledgments are most frequent from authors with U.S. affiliations, followed by China and India in the fields of Biomedical, Clinical Sciences, and Information and Computing Sciences in avenues like "The Lancet Digital Health" and "bioRxiv."

#### 2.1.5. Exams

Interactive Professional exam studies contend that ChatGPT consistently outperforms Bard, and 50–75% of fundamental questions of SAT were answered incorrectly by Bard [77]. In clinical medicine, Bard shows a higher probability of generating hallucinations [78], while ChatGPT provides high-quality and relevant responses to questions about postpartum depression and imaging [79]. Campello et al. [80] compared the performance of Bing, Bard, ChatGPT, and Quora Poe in an intelligence test in Brazil and contend that ChatGPT and Bing achieve the highest scores and possess the intellectual ability to replace humans in high-skilled jobs.

#### 2.1.6. Human resources settings

Though GAI tools have produced impressive results in various areas, their potential in HR settings remains largely unexplored. The primary industries exposed to advances in GAI tools are commodities and investments, securities, and legal services [81]. Since the exposure to GAI tools and occupations' mean wages exhibit a statistically significant correlation [81], it contradicts the notion that GAI would first impact monotonous and high-risk jobs. Assessing human resource management from the point of GAI tools enables one to discover the potential ethical risks of GAI tools [82] and facilitate the development of more trustworthy GAI tools. It would help determine the perception of humans and understand their modes of thinking, response motivation, and communication patterns [51].

GAI tools like ChatGPT and Bard can provide employees with human capital development resources and provide personalized feedback. Hence, the purpose of our study is to evaluate the performance of ChatGPT and Bard in their HR literacy, which can provide insights into their strengths and weaknesses and the roles they are most suited to take up in HR. Our methodology of using HR certification to assess HR literacy remains largely unexplored since HR academicians tend to focus on issues of less interest to practitioners, and consequently, there are limited efforts to collect data on certifications [83]. We address this gap by comparing the HR literacy levels of two Generative Artificial Intelligence (GAI) tools, ChatGPT and Bard, using a dataset of an HR certification.

### 2.2. Society for Human Resource Management (SHRM) certification

SHRM, the most extensive professional human resource association founded in 1948, is a non-profit professional membership organization (<https://www.shrm.org/>). It has over 300,000 members devoted to promoting, educating, and connecting HR professionals in the workplace [84]. Individuals in Human Resource Management acquire certifications to establish credibility and competence [85–87]. The SHRM Certified Professional (SHRM-CP) deals with topics like implementing HR policies, data collection, compliance, and discipline. The SHRM-CP comprises 134 questions, segregated into situation-based and knowledge-based categories. Each category is timed independently and has to be completed within 3 h and 40 min.

## 3. Research framework

Our research focuses on evaluating the quality of responses generated by GAI tools like ChatGPT and Bard. The research question is motivated by the growing importance of these tools in various applications, where accurate and reliable responses are crucial. Understanding and improving the quality of their responses is essential for ensuring their responsible and effective utilization.

- **Inconsistent accuracy:** GAI tools ChatGPT and Bard can exhibit inconsistencies in their response accuracy, making it difficult to rely on them for critical tasks. This research aims to identify factors influencing accuracy and develop methods for improving it.
- **Limited comprehensiveness:** GAI tools may sometimes provide incomplete or superficial responses, lacking sufficient depth and detail. This research seeks to assess the comprehensiveness of responses and explore ways to ensure they cover all relevant information.
- **Ambiguity and redundancy:** GAI tools responses can be unclear or convoluted, making them difficult to understand. This research investigates methods to enhance clarity and conciseness, leading to unambiguous and direct responses.

The rapid integration of GAI tools into various domains necessitates a comprehensive understanding of their response quality. Numerous tests exist across different domains, yet the distinct features of each field, like human resources, require specialized studies. This is due to the fact that GAI tools might display varied behaviors and susceptibilities in specific domain contexts, which cannot be

generalized from standard tests. The noted variances in how ChatGPT and Bard manage sensitive issues are indicative of their inherent ethical and philosophical coding, which holds importance. Comprehending these variations is crucial as these tools become more integrated into society and as we face ethical challenges. This research is crucial for enabling responsible and effective utilization of GAI tools in domains such as education, healthcare, and customer service. By identifying and addressing the limitations in response quality, this research paves the way for improved GAI tools' performance and increased trust in their generated outputs.

**Table 1** serves as a comprehensive research framework designed for the empirical evaluation of ChatGPT and Bard. The framework employs a multi-dimensional approach to investigate the performance of the two tools on a dataset encompassing questions from the Human Resources Management (HRM) domain. The metrics are organized into Descriptive Analysis, Response Analysis, Statistical Analysis, Impact of Confirmation Queries, Cosine Similarity Analysis, Readability Analysis, and Comparative Analysis of responses.

**Research Problem 1: Variability in Response Quality.** This research aims to quantify and analyze the performance variability in AI responses. By utilizing descriptive analysis and t-tests, the study identifies the extent of variability and performance differences between models.

**Research Problem 2: Inconsistency in Response Structure and Content.** The study uses cosine similarity analysis to evaluate how consistent the GAI tools are in their response structure and content, addressing concerns about reliability.

**Research Problem 3: Readability and Accessibility of AI-generated Content.** Using Flesch metrics, the research assesses the readability and complexity of language used by GAI tools, which is crucial for ensuring that the content is accessible to a broad audience.

**Research Problem 4: Comparative Analysis of Responses.** Here we looked at several scenarios comparing ChatGPT and Bard's responses. For each of the scenarios, like both responses were correct or incorrect, only one of them is correct etc, we performed further analysis.

### 3.1. Study design

We used ChatGPT and Bard to evaluate their performance in responding to MCQ questions from a dataset (n = 134 MCQs) prepared by the Society for Human Resource Management (SHRM). Our study was conducted using the GPT-4 (<https://chat.openai.com/>) and Bard (<https://bard.google.com/chat>, updated July 13, 2023) from August 15 to 21, 2023.

Data Collection consisted of the following steps.

1. Question Input: Each question from the dataset was systematically input into ChatGPT and Bard.
2. Initial Response Capture: Responses from both were recorded.
3. Confirmation Query: If the initial response from either ChatGPT or Bard did not match the correct answer, a follow-up query, "Are you sure?" was posed to the respective tool.
4. Second Response Capture: The subsequent response was recorded after the confirmation query.
5. Explanation Capture: Regardless of the correctness of the second response, an explanation for the given response was recorded and analyzed by experts for three measures (Accuracy, Relevance, and Clarity)

### 3.2. Descriptive and Response Analysis

**Table 2** shows how we track the response accuracy from both tools.

- "Both Correct" means that ChatGPT and Bard answered the question correctly.
- "Bard Only Correct" means that Bard answered the question correctly, while ChatGPT did not.
- "ChatGPT Only Correct" means that ChatGPT answered the question correctly, while Bard did not.
- "Both Wrong" means that neither ChatGPT nor Bard answered the question correctly.

**Table 1**  
Research framework to evaluate ChatGPT and Bard.

Measure	Metric	Interpretation
Descriptive Analysis	Average and distribution of scores	Higher averages indicate better performance; distribution helps understand variability in responses.
Response Analysis	Accuracy Relevance Clarity	How correct and truthful the response is. Broader coverage of relevant information in the response. Clearer responses without ambiguity.
Statistical Analysis	t-test results for each measure	Significant differences show that one tool may outperform the other in certain measures.
Impact of Confirmation Queries	The difference in accuracy when the query 'Are you sure?' is posed	A positive difference suggests that the confirmation query improves accuracy.
Cosine Similarity Analysis	Cosine similarity values between ChatGPT and Bard's responses	Higher values indicate more similarity in response structure and content, which might indicate consistency.
Readability Analysis	Flesch Readability Ease Flesch-Kincaid Grade Level	Higher scores indicate easier readability. Lower scores indicate simpler language, targeting a younger audience.
Comparative Analysis of responses	Select a sample of questions.	Which responses are correct, incorrect, only one tool has answered correctly, no answer, etc.?

**Table 2**  
Accuracy analysis.

	ChatGPT Correct	ChatGPT Wrong
Bard Correct	Both Correct	Bard Only Correct
Bard Wrong	ChatGPT Only Correct	Both Wrong

We also tracked questions that either ChatGPT or Bard chose not to select any of the given answer choices.

We are proposing three measures, Accuracy, Relevance, and Clarity, to categorize the responses to questions. Each response is rated on a 5-point Likert scale.

**Measure 1:** Accuracy of the response, considering how correct and truthful the response is.

1. Completely Inaccurate: The response is entirely incorrect or false.
2. Mostly Inaccurate: The response contains some truth but is largely incorrect.
3. Somewhat Accurate: The response is a mixture of correct and incorrect information.
4. Mostly Accurate: The response is largely correct but may contain minor inaccuracies.
5. Completely Accurate: The response is entirely correct and truthful.

**Measure 2:** Relevance of the response, considering how closely it aligns with the question or prompt.

1. Completely Irrelevant: The response does not address the question or prompt at all.
2. Mostly Irrelevant: The response somewhat addresses the question but is largely off-topic.
3. Somewhat Relevant: The response addresses the question to some extent but could be more focused.
4. Mostly Relevant: The response is on-topic and closely aligns with the question, with minor deviations.
5. Completely Relevant: The response directly addresses the question or prompt and stays on-topic.

**Measure 3:** Clarity of the response, considering how clearly and understandably the response is communicated.

1. Very Confusing: The response is largely incoherent and difficult to follow.
2. Mostly Confusing: The response has some clear elements but is overall hard to understand.
3. Moderately Clear: The response is understandable but could be clearer or more concise.
4. Mostly Clear: The response is clear, with only minor ambiguities or complexities.
5. Completely Clear: The response is articulated in a perfectly clear and coherent manner.

Two independent expert faculty members from the Business School at our University who regularly teach students graduate-level courses in Human Resources Management were assigned to read each explanation text from ChatGPT and Bard and assign a score of 1–5 for each of the three measures. To ensure the reliability and consistency of the scoring process, we used a third expert faculty member as an adjudicator. This expert member was used whenever the first two members disagreed on a score. The third expert's role was to review the explanation text and decide which was considered final.

### 3.3. Statistical analysis

Paired sample *t*-tests were employed to assess the differences in scores between ChatGPT and Bard for each measure:

1. Comprehensiveness *t*-test: A paired *t*-test was conducted to compare the accuracy scores of ChatGPT and Bard for each question.
2. Clarity *t*-test: Similarly, another paired *t*-test was performed to contrast the relevance scores of the two GenAI tools for every question.
3. Conciseness *t*-test: A third paired *t*-test was carried out to compare the clarity scores of ChatGPT and Bard for each question.

Interpretation of Results: Significance was determined at the 0.05 level. For each *t*-test, a *p*-value of less than 0.05 would suggest a statistically significant score difference between the two GenAI tools for that metric. The mean difference would indicate the direction of this difference (i.e., which model performed better).

### 3.4. Cosine similarity analysis

Cosine similarity is a measure of the similarity between two texts. In the context of comparing the responses of ChatGPT and Bard, cosine similarity can be used to measure the similarity between the text embeddings of the two responses. A study by Borji et al. [88] used cosine similarity to compare the responses of four different large language models (LLMs), including ChatGPT and Bard. The study found that cosine similarity was a reliable measure for comparing the responses of LLMs and that it could be used to identify differences in the performance of the different models.

Calculating the cosine similarity between the answers given by ChatGPT and Bard for the same question can provide insights into

how similarly the two models are thinking. If the cosine similarity is high, it suggests that both models are using similar terms and concepts in their answers. If it's low, they are approaching the question differently. One advantage of using cosine similarity is that it is a relatively simple and efficient measure to compute. Another advantage is that it is not affected by the length of the text embeddings. This makes it a suitable measure for comparing the responses of ChatGPT and Bard, which may have different lengths. A publicly available tool was used to calculate the Cosine similarity values (<https://tilores.io/cosine-similarity-online-tool>). Response for each question by ChatGPT and Bard was copied into the tool, and the resulting similarity value was noted.

### 3.5. Readability Analysis

Calculating the Flesch Readability Ease score and the Flesch-Kincaid Grade Level score for each response from ChatGPT and Bard can be a valuable way to compare how the two models perform regarding readability. These scores provide quantitative measures of how easy reading and understanding a given text is.

**Flesch Readability Ease score:** Measures the reading ease of a text, with higher scores indicating easier readability. Comparing these scores between the two models can provide insights into which model produces text that is easier for a general audience to read.

**Flesch-Kincaid Grade Level score:** Estimates the U.S. educational grade level required to understand the text. Comparing these scores can give you an idea of the complexity of the responses generated by each model.

In a study by Patnaik and Hoffmann [53], the readability of the responses of ChatGPT and Bard was compared using the Flesch Readability Ease score and the Flesch-Kincaid Grade Level score. The study found that the responses of Bard had a higher Flesch Readability Ease score and a lower Flesch-Kincaid Grade Level score than the responses of ChatGPT. This suggests that the responses of Bard were easier to read and understand than the responses of ChatGPT.

A publicly available tool calculates the readability scores (<https://goodcalculators.com/flesch-kincaid-calculator/>). Response for each question by ChatGPT and Bard was copied into the tool, and resulting values were interpreted.

## 4. Results

### 4.1. Descriptive analysis

Table 3 offers a detailed comparative analysis of the accuracy between ChatGPT and Bard's answers based on 134 questions from the SHRM dataset. In scenarios where both Bard and ChatGPT provided correct answers, Bard was correct 99 times. When Bard was correct and ChatGPT was incorrect, this occurred in 12 instances. On the other hand, Bard was incorrect, while ChatGPT was correct in 15 cases. Both tools were incorrect in 8 instances. This results in an overall accuracy rate of 84.3% for ChatGPT (114 correct answers out of 134) and 82.8% for Bard (111 correct answers out of 134).

An intriguing aspect of this comparison is revealed through the nuanced discrepancies in cases where one model was correct while the other was not, as well as situations where both models made the same mistakes. These findings illuminate the distinct strengths and weaknesses of each GAI tool and indicate potential variances in their design or training data, which could influence their performance in specific domains.

From Table 4, based on the Paired Samples *t*-test results, the mean accuracy for ChatGPT (4.86) and Bard (4.80) suggests a minimal difference in performance between the two systems. The variance indicates that ChatGPT's scores are more tightly clustered around the mean compared to the Bard, implying more consistent performance. The Pearson correlation of approximately 0.23 shows a weak positive association between the accuracies of the two systems. Crucially, the *p*-values for both one-tailed (0.152) and two-tailed (0.303) tests exceed the conventional alpha level of 0.05, indicating that the observed differences in accuracy are not statistically significant. In summary, the statistical analysis does not provide sufficient evidence to conclude that there is a meaningful difference in the accuracy of ChatGPT and Bard for the dataset under examination.

As seen in Table 5, the Paired Samples *t*-test results for the "Relevance" metric provide several notable insights. Both ChatGPT and Bard display remarkably similar mean relevance scores - 4.93 for ChatGPT and 4.92 for Bard, indicating a negligible difference in performance on this measure. Variance in relevance for ChatGPT (0.085) is slightly lower than for Bard (0.121), suggesting that ChatGPT's performance may be marginally more consistent in terms of relevance. The Pearson correlation is extremely low at approximately 0.013, suggesting almost no linear relationship between the relevance scores of the two systems. Regarding the test statistics, the *t*-value is 0.192, which is much lower than the critical *t*-values for both one-tailed (1.656) and two-tailed (1.978) tests. The *p*-values for one-tailed (0.424) and two-tailed (0.848) tests are significantly greater than the conventional significance level of 0.05. This implies that the observed minute difference in relevance between the two systems is statistically insignificant. In summary, the statistical evidence does not support a significant difference in the relevance performance between ChatGPT and Bard based on this dataset.

In the Paired Samples *t*-test results as shown in Table 6 concerning the "Clarity" metric, the average clarity score for ChatGPT stands

**Table 3**  
Accuracy of responses.

	ChatGPT Correct	ChatGPT Incorrect
<b>Bard Correct</b>	99	12
<b>Bard Incorrect</b>	15	8

**Table 4**  
t-test: Paired Two Sample for Means (Accuracy).

	Accuracy ChatGPT	Accuracy Bard
Mean	4.858209	4.798507
Variance	0.197789	0.372629
Observations	134	134
Pearson Correlation	0.226317	
Df	133	
t Stat	1.033055	
P (T ≤ t) one-tail	0.151727	
t Critical one-tail	1.656391	
P (T ≤ t) two-tail	0.303453	
t Critical two-tail	1.977961	

**Table 5**  
t-test: Paired Two Sample for Means (Relevance).

	Relevance ChatGPT	Relevance Bard
Mean	4.925373	4.91791
Variance	0.084615	0.12103
Observations	134	134
Pearson Correlation	0.013307	
df	133	
t Stat	0.191757	
P (T ≤ t) one-tail	0.424112	
t Critical one-tail	1.656391	
P (T ≤ t) two-tail	0.848225	
t Critical two-tail	1.977961	

at approximately 4.86, whereas for Bard, it is around 4.62. This suggests a more noticeable difference in performance compared to the “Accuracy” and “Relevance” metrics previously discussed. ChatGPT also demonstrates a lower variance (0.138) relative to Bard (0.268), which could indicate more consistent clarity scores. The Pearson correlation between the two sets of clarity scores is 0.187, denoting a weak positive relationship. Importantly, the t-value is substantially high at 4.79, which far exceeds the critical t-values for both one-tailed (1.656) and two-tailed (1.978) tests. The p-values are remarkably low - 2.22E-06 for the one-tailed and 4.45E-06 for the two-tailed tests—much below the conventional alpha level of 0.05. Consequently, we can reject the null hypothesis, concluding that there is a statistically significant difference in clarity between ChatGPT and Bard. The statistical results strongly indicate that ChatGPT performs significantly better than Bard in terms of clarity based on the examined dataset.

#### 4.2. Impact of confirmation query

Table 7 illustrates the impact of a confirmation query, specifically “Are you sure?” on the responses of Bard and ChatGPT.

For Bard, only one response was correct after the “Are you sure?” query, whereas 23 responses were incorrect. On the other hand, ChatGPT had zero correct responses following the query but also fewer incorrect responses, with a count of 16.

This data suggests that the inclusion of a confirmation query (“Are you sure?”) does not seem to improve the accuracy of either Bard or ChatGPT. Specifically, Bard appears to be notably inaccurate after the confirmation query, and ChatGPT does not achieve any correct responses post-query but has fewer incorrect ones compared to Bard.

The implications of these findings may point to limitations in the ability of GAI tools to correctly adjust their responses when prompted for verification, thus raising questions about the efficacy of using a confirmation query as a mechanism to improve response accuracy.

**Table 6**  
t-test: Paired Two Sample for Means (Clarity).

	Clarity ChatGPT	Clarity Bard
Mean	4.858208955	4.619403
Variance	0.137638873	0.267591
Observations	134	134
Pearson Correlation	0.186826476	
df	133	
t Statv	4.786702082	
P (T ≤ t) one-tail	2.22383E-06	
t Critical one-tail	1.656391244	
P (T ≤ t) two-tail	4.44766E-06	
t Critical two-tail	1.977961264	



**Table 7**  
Impact of Confirmation Query – Are you sure?

	Correct after Are you sure	Incorrect after Are you sure
<b>Bard</b>	1	23
<b>ChatGPT</b>	0	16

#### 4.3. Cosine similarity analysis

In the analysis of 134 questions, observable changes in cosine similarity scores emerged from the dataset. A small subset, comprising 12 questions, displayed a cosine similarity score below 60, implying a low level of congruence between responses from ChatGPT and Bard. In contrast, the vast majority, consisting of 113 questions, exhibited scores ranging from 60 to 80, denoting a moderate to high degree of similarity between the two systems. Exceptionally, three questions had scores surpassing 80, signifying an almost identical response generated by both tools. Intriguingly, for six questions, Bard was unable to generate a response, rendering the computation of a similarity score unfeasible. Organizations and developers responsible for GenAI tools may adopt varying approaches to address ethical concerns or potential misuse. OpenAI, for instance, has continually fine-tuned models like ChatGPT to strike a balance between utility and safety, whereas Bard might have been engineered with more stringent safeguards against misuse.

#### 4.4. Readability Analysis

In the analysis of readability metrics, the Flesch-Kincaid Grade Level score and the Flesch Readability Ease score offer contrasting insights into the language complexity and readability of text generated by ChatGPT. On the Flesch Readability Ease score, ChatGPT registers at 31.491, while Bard scores higher at 41.608 (Table 8). This scale generally ranges from 0 to 100, with lower scores indicating more difficult text and higher scores indicating easier text. Hence, according to this metric, Bard's output is more straightforward to read than ChatGPT's. This suggests that text generated by ChatGPT may require a higher educational level for comprehension, implying a more complex linguistic structure. Conversely, Bard's text appears to be targeted at a slightly lower educational level.

The diverging scores on Flesch-Kincaid Grade Level and Flesch Readability Ease suggest that while ChatGPT's output may employ a more complex lexicon or sentence structure, Bard's output is geared towards more comfortable readability. Therefore, Cosine Similarity could reveal whether these disparities in readability and complexity translate into significant differences in the actual word and phrase usage between the two systems.

#### 4.5. Comparative Analysis of Responses

The differing responses between ChatGPT and Bard when fielding questions related to human resource management may be influenced by various factors:

- **Data Training:** The functioning of these AI models is predominantly influenced by their training data. Distinctive datasets or fine-tuning approaches could lead ChatGPT or Bard to offer more guarded responses.
- **User Feedback:** Modifications to the model can also come from user input. If developers note that the system is being exploited or misused in some manner, they may fine-tune the model's parameters or guidelines to make it more cautious.

In addition to these main factors, several other considerations could affect their responses. For example, the specific algorithms, filtering techniques, and any built-in rules that are part of each system can differ and thus influence the model's behavior. Furthermore, the models' generalization capabilities from their training data could also vary. When posed with politically sensitive queries, Bard takes an opinionated stand, while ChatGPT tends to be neutral [69]. If, during training, ChatGPT or Bard were exposed to instances where human resource management questions were linked to contentious or unethical situations, the model might become more reserved in providing detailed responses in this domain. Thus, a combination of factors can contribute to the variations in behavior observed between ChatGPT and Bard in the realm of human resource management.

We have looked at several scenarios comparing ChatGPT and Bard's responses (Table 9).

Comparative analysis results.

**Table 8**  
Readability analysis.

GenAI tool	Flesch-Kincaid Grade Level score	Flesch Readability Ease score	Estimated reading level
ChatGPT	12.505	31.491	College
Bard	11.469	41.608	College

## 5. Discussion on comparative analysis

The comparative analysis of Bard and ChatGPT's efficacy in providing strategic HR solutions may be conducted by examining the accuracy of the responses shown in [Tables 11–15](#) and the underlying HR strategy articulated within these responses. In our research, we consider both quantitative and qualitative data, encompassing metrics such as the proportion of accurate and inaccurate responses and the comprehensiveness exhibited in the problem-solving procedure.

[Table 11](#) illustrates that ChatGPT and Bard have adopted a quantitative, data-centric approach, indicating a proclivity towards utilizing empirical data in developing human resources strategies and decision-making. This aligns with contemporary human resources practices, which emphasize utilizing large-scale data and data analytics. According to the findings presented in [Table 11](#), both ChatGPT and Bard recommend selecting option "c) Monthly turnover levels" as the preferred choice for Question 12. Based on the information shown in the table, it can be concluded that option "a) Exit interviews" is the correct choice for the SHRM test.

The underlying concept of ChatGPT and Bard's response was to acquire a quantitative comprehension of staff turnover by analyzing monthly attrition rates. Rather than relying solely on post-departure exit interviews, human resources practitioners should utilize timely and comprehensive data to discern patterns in employee departures and modify their strategies accordingly. In contrast, Option (a) exit interviews indicate that the organization is endeavoring to get further insights into the factors contributing to employee departures. This assertion is substantiated by the suitable response in the SHRM evaluation, as depicted in the table provided. Merely examining turnover rates in isolation may yield insignificant findings for the company. When undertaking exit interviews, it is possible to acquire qualitative data that aids in elucidating the factors contributing to an employee's departure. After collecting this data, it can be employed to develop more efficient approaches to maintain existing personnel.

According to [Tables 12](#) and it is evident that ChatGPT exhibits specific inconsistencies that possess the potential to give rise to challenges within a strategic human resources framework, whereby consistency has paramount importance. Bard advocates for human resource management practices, prioritizing preventive measures and forward-thinking approaches. According to the findings presented in [Tables 12](#) and it can be inferred that option d) represents the most suitable course of action in response to issue 8. This implies that the human resources department should guide the organization, emphasizing the need to acknowledge the concerns raised and recommending a cautious approach in terms of immediate action. Although ChatGPT first selected option "c" before transitioning to option "a," Bard ultimately opted for the option "a." A pragmatic method can be applied to elucidate the rationale behind the organization's adoption of this stance, as seen by the correct response (d) presented in the table. This approach upholds the organization's existing position about the rigorous dress code policy while also considering the apprehensions expressed by the sales representative. One possible interpretation of the choice could be a managerial trade-off between the desire for predictability and the need for flexibility. Addressing criticism while maintaining the potential for further investigation demonstrates a prudent approach. Furthermore, it postpones the commencement of an investigation, allowing the organization to maintain compliance with its current protocol while remaining open to suggestions.

The response provided by SHRM presents a viewpoint that proposes a potential compromise, indicating that although the concerns are valid, they do not necessitate an immediate investigation. This approach makes it possible to maintain the policy's implementation while simultaneously adapting it in response to our deepening understanding of its far-reaching consequences.

The depiction of ChatGPT is founded upon the prevailing depiction of the contemporary cooperative arrangement, as outlined in [Table 13](#). This strategy can be advantageous for HR professionals because it emphasizes the importance of consistent relationships and promotes open communication. According to the data presented in [Table 13](#), the response of ChatGPT is option a) Schedule weekly cross-departmental manager meetings to discuss department functions and challenges. In contrast, the response of Bard is option c) Temporarily transfer department managers across departments to gain experience working with different groups. At the same time, ChatGPT gives a more transactional answer, which is more practical and involves less conflict. Hence, we contend that ChatGPT would be more beneficial to augment the transactional processes in the HR function. In the actual workplace, GAI tools would be more helpful to augment transactional roles than strategic roles of the HR function.

According to [Table 14](#), both Bard and ChatGPT have demonstrated a comprehensive comprehension of international management practices. This particular skill serves as a valuable asset in enhancing the adaptability required in modern human resources contexts and is crucial for effectively addressing the complexities inherent in global human resources management. The answer provided for question 44 in [Table 14](#) is "a) Ethnocentric." The initial selection of "d) Polycentric" by ChatGPT was evidently erroneous. The Bard responded promptly once more on this occasion.

The SHRM's choice promotes an ethnocentric approach to global management by advocating for the importation of American methods and legal frameworks into other markets, disregarding other nations' cultural and economic realities. The ethnocentric

**Table 9**

Comparative analysis scenarios comparing ChatGPT and Bard's responses.

Scenarios	Table Number
Both ChatGPT and Bard gave correct answers.	<a href="#">Table 10</a>
Both ChatGPT and Bard gave wrong answers but agreed on the wrong answer	<a href="#">Table 11</a>
Both ChatGPT and Bard gave wrong answers but did not agree on the answer	<a href="#">Table 12</a>
ChatGPT's answer is correct, and Bard's answer is wrong.	<a href="#">Table 13</a>
ChatGPT's answer is wrong, and Bard's answer is correct	<a href="#">Table 14</a>
Bard chose not to answer, citing language model constraints	<a href="#">Table 15</a>

**Table 10**

Both ChatGPT and Bard gave correct answers.

Q	Question	Correct Answer	ChatGPT's Explanation	Bard's Explanation
2	<p>Mark the correct option. A company has been giving its employees automatic pay increases annually. Recently, however, production has decreased, resulting in less profit. The president tasks HR with implementing a new goal-oriented pay plan to incentivize employees to increase production. HR designs a pay-for-performance plan and trains supervisors. HR also sends an e-mail to all employees explaining the change, referring employees with questions to their supervisor. As part of the new program, goal development is handled by supervisors. However, some supervisors feel that the change in pay philosophy is unnecessary and tell employees that the new pay plan will not work. Twelve months pass, and, upon review of performance evaluations, HR finds that some appraisals lack goals and do not objectively measure employees' work performance. The results are recommended pay increases that are subjective and unsubstantiated by metrics. Rumors that pay increases will be provided to employees who are favorites begin to impact morale, and work production continues to decline. Leadership is disappointed with the results of using a pay-for-performance system. Which first step should the HR manager take to change the current perceptions of HR with the company's managers?</p> <p>a) Recommend reinstatement of the prior pay plan with minimal communication. b) Recommend reinstatement of the prior pay plan with minimal communication. c) Notify all employees about the retraction of the new pay plan to avoid employee complaints and supervisor dissatisfaction. d) Communicate the organization's mission, vision, and strategy and HR's role in support of the organization's and employees' success.</p>	<p><b>d) Communicate the organization's mission, vision, and strategy and HR's role in support of the organization's and employees' success.</b></p>	<p>The situation described highlights issues related to communication, alignment with organizational strategy, and buy-in from supervisors. The problem does not necessarily lie in the pay-for-performance plan itself, but rather in its implementation, understanding, and acceptance. Given the options provided: a) and b) are the same and both suggest simply reverting to the old pay plan without addressing the underlying problems. c) Focuses on retraction of the new plan, which again fails to address the underlying issues and misalignment. d) Communicate the organization's mission, vision, and strategy and HR's role in support of the organization's and employees' success. This option seems to focus on aligning managers and employees with the organizational strategy and clarifying HR's role in supporting success. <b>It seems to be the most logical first step in addressing the core issues.</b></p> <p>So, the correct answer would be: d) Communicate the organization's mission, vision, and strategy and HR's role in support of the organization's and employees' success.</p>	<p><b>The answer is d) Communicate the organization's mission, vision, and strategy and HR's role in support of the organization's and employees' success.</b></p> <p>The HR manager needs to first address the root cause of the problem, which is the lack of trust between employees and supervisors. This can be done by communicating the organization's mission, vision, and strategy, and HR's role in support of the organization's and employees' success. This will help to clarify the purpose of the pay-for-performance plan and why it is important for the company.</p> <p>Once the trust has been rebuilt, the HR manager can then address the specific problems with the pay-for-performance plan. This may involve retraining supervisors on how to develop goals and how to write performance evaluations. It may also involve revising the pay-for-performance plan to make it more objective and transparent.</p> <p>Recommending reinstatement of the prior pay plan with minimal communication (option a) is not a good idea. This would only serve to confirm employee suspicions that HR is not acting in their best interests.</p> <p>Notifying all employees about the retraction of the new pay plan to avoid employee complaints and supervisor dissatisfaction (option c) is also not a good idea. This would only serve to further damage employee morale and trust in HR.</p> <p>Therefore, the best first step for the HR manager is to communicate the organization's mission, vision, and strategy and HR's role in support of the organization's and employees' success. This will help to rebuild trust and lay the foundation for addressing the specific problems with the pay-for-performance plan.</p> <p>Here are some additional things that the HR manager can do to change the current perceptions of HR with the company's managers:</p> <p>Meet with managers one-on-one to discuss their concerns about the pay-for-performance plan. Set up a task force of managers and employees to review the pay-for-performance plan and make recommendations for improvement. Conduct a survey of employees to get</p>

(continued on next page)

Table 10 (continued)

Q	Question	Correct Answer	ChatGPT's Explanation	Bard's Explanation
				<p>their feedback on the pay-for-performance plan.</p> <p>Hold town hall meetings to discuss the pay-for-performance plan with employees.</p> <p>By taking these steps, the HR manager can demonstrate that they are committed to working with employees and managers to create a fair and equitable pay system. This will help to rebuild trust and improve the perception of HR in the company.</p>

approach aligns with a global pay and benefits plan that overlooks the diversity of local cultures. The policymaking process prioritizes the preferences and norms of the home country, often overlooking the distinct requirements and preferences of diverse localities.

Table 15 illustrates the divergent reaction patterns of Bard and ChatGPT, potentially indicating disparities in their fundamental training, programming, and methods towards handling sensitive and crucial human resources situations.

ChatGPT demonstrates a comprehensive engagement with each scenario, providing in-depth analysis and displaying a proactive attitude toward addressing complex human resources challenges. In contrast, Bard is more cautious, refraining from offering elaborate explanations. In HR scenarios characterized by intricate circumstances involving legal and ethical considerations, a more prudent strategy may be adopted to mitigate the potential for providing inaccurate or excessively assertive guidance.

Upon examining specific inquiries that incorporate the term “strategy,” namely questions 22 and 11, it becomes apparent that Bard refrains from responding. This abstention potentially signifies a predetermined tendency to refrain from participating in strategic discussions without a comprehensive understanding and contextual awareness. This choice points towards the fact that Bard has more safeguards built into its design than ChatGPT against misuse in the HR function. Bard refrains from providing answers to inquiries numbered 89, 110, and 118 due to their nature, including potentially sensitive matters related to human resources, specifically about pay discrepancies and safeguards for those reporting misconduct. These particular queries require a comprehensive knowledge of the circumstances at hand. We reason that while ChatGPT emphasizes the transactional response focusing on technical aspects of the job and completion of the task at hand, Bard takes a cautious stand and expresses an inability to answer the questions that require more contextual data before making the right decisions.

Upon comparing the arguments provided by the two GAI tools, it appears that Bard presents a more comprehensive and detailed elucidation of the reasons for the selection or rejection of each alternative (Table 16). This strategy considers a more significant amount of information before decision-making, resulting in a more comprehensive study of each issue. The top selections of ChatGPT were supported by robust statistical evidence and a sound rationale. However, the specific criteria for removing individual remarks were not readily apparent.

Human Capital theory [89] emphasizes the role of individual and unit-level human capital – the knowledge, skills, and abilities. The Human Resource function in organizations deals with the development of employee expertise to improve the performance of individuals, teams, and organizations [90]. In the current organizational context, there is a growing demand for individualized development possibilities that could be accessed at any time and place to suit individual needs. Our findings that both Bard and ChatGPT possess notable strengths and weaknesses when it comes to addressing strategic and transactional human resources (HR) concerns could provide insights into the roles they are most suited to take up advancing the Human Capital in organizations. We find that while Bard demonstrates proficiency in strategic thinking, their propensity to engage in excessive contemplation can lead to reticence or even induce hallucinatory experiences.

In contrast, ChatGPT exhibits commendable efficacy in providing prompt and pertinent replies. Our study provides evidence that the HR literacy levels of ChatGPT are higher than that of Bard, and we conclude that ChatGPT can serve as a robotic advisor [91,92] for transactional HR roles. ChatGPT, as the robotic advisor, can be used to provide employees with access to develop human capital and provide personalized feedback for skill development.

### 5.1. GAI tools strengths and weakness

One notable observation regarding the Bard system is its tendency to exhibit delayed response times, presumably due to the cognitive process of thoughtful consideration prior to providing an answer. Providing immediate responses can be challenging due to the presence of a strategic inclination. It appears that the Bard platform has been specifically developed to enhance the capacity for comprehensive analysis of various choices, hence offering potential benefits for implementing strategic human resources planning effectively. This aligns with prior research that has shown the benefits of adopting strategic HRM practices inside organizations [8]. Nevertheless, this inclination may occasionally hinder Bard's ability to provide direct responses to practical and evident inquiries.

Given the apparent propensity of Bard to increase the probability of experiencing hallucinations, it is imperative to implement additional measures to prevent its misuse. While the inclination towards protection is undeniably crucial in upholding moral standards, there exist circumstances in which its utility may be diminished. Given the extensive research conducted on AI hallucinations,

**Table 11**  
Both ChatGPT and Bard gave wrong answers but agreed on the wrong answer.

Q	Question	Correct Answer	ChatGPT Explanation	Are you sure?	Bard's Explanation	Are you Sure?
12	Mark the correct option. A new competitor entered the market and attracted many of a manufacturing organization's employees with higher hourly wages. The organization's HR team has revised the staffing plan. Which data could alert HR to the need for further revisions? a) Exit interviews b) Stay interviews c) Monthly turnover levels d) Absenteeism reports	<b>a) Exit interviews</b>	<p>In the scenario described, a manufacturing organization is losing employees to a competitor offering higher hourly wages, and the HR team has revised the staffing plan. The question asks which data could alert HR to the need for further revisions.</p> <p>The best metric to gauge whether the revised staffing plan is working would be to look at how many employees continue to leave the organization. If the turnover remains high, it might signal that the revisions to the staffing plan are not sufficient to retain employees.</p> <p>Therefore, the correct option is: <b>c) Monthly turnover levels.</b></p> <p>This metric would directly reflect the ongoing trend in employee departures and help HR understand if further adjustments are necessary.</p>	<p>Yes, I believe the correct option is indeed <b>c) Monthly turnover levels.</b></p> <p>Here's a brief explanation of why the other options might not be as directly relevant:</p> <p>a) Exit interviews: While these can provide valuable insights into why employees are leaving, they may not provide ongoing, quantitative data to signal the need for further staffing plan revisions.</p> <p>b) Stay interviews: These interviews with current employees can help understand what keeps them with the company but might not directly alert HR to the need for revisions in the staffing plan.</p> <p>d) Absenteeism reports: While this might indicate dissatisfaction among employees, it's not as directly tied to the specific problem of losing employees to a competitor with higher wages.</p> <p>Option c) Monthly turnover levels would provide a regular, quantitative measurement of how many employees are leaving the organization, directly reflecting the issue at hand and alerting HR to the need for further revisions if the turnover rate remains high.</p>	<p><b>The correct answer is c) Monthly turnover levels.</b></p> <p>Monthly turnover levels can alert HR to the need for further revisions to the staffing plan. If the monthly turnover levels are high, it could be a sign that the new competitor is still attracting employees with higher hourly wages.</p> <p>Exit interviews and stay interviews can also be helpful in identifying factors that are contributing to turnover. However, they are not as timely as monthly turnover levels. Absenteeism reports can also be helpful in identifying factors that are contributing to turnover. However, they are not as specific as monthly turnover levels.</p> <p>Therefore, the best data to alert HR to the need for further revisions to the staffing plan is monthly turnover levels.</p> <p>Here are some other factors that HR can consider when monitoring turnover levels:</p> <p>The industry average turnover rate. The company's historical turnover rate. The reasons for turnover. The positions that are most affected by turnover. By monitoring these factors, HR can get a better understanding of the reasons for turnover and make adjustments to the staffing plan as needed.</p>	<p>Yes, I am sure.</p> <p>Exit interviews and stay interviews are helpful in identifying factors that are contributing to turnover, but they are not as timely as monthly turnover levels. Absenteeism reports can also be helpful in identifying factors that are contributing to turnover, but they are not as specific as monthly turnover levels.</p> <p>Monthly turnover levels are the most timely and specific data that HR can use to identify the need for further revisions to the staffing plan. If the monthly turnover levels are high, it could be a sign that the new competitor is still attracting employees with higher hourly wages.</p> <p>I hope this helps. Is there anything else I can help you with?</p>

including an examination of their associated risks and ethical implications, the existing body of literature strongly advocates for implementing techniques to reduce these issues [78].

The responses generated by ChatGPT are characterized by their transactional nature and ease of implementation, hence circumventing strategic intricacies. The enhanced efficacy of ChatGPT in operational HR functions that require prompt and pragmatic resolutions may be accompanied by a diminished understanding of abstract strategic concepts. According to Ulrich [93], the proper

**Table 12**  
Both ChatGPT and Bard gave wrong answers but did not agree on the answer.

Q	Question	Correct Answer	ChatGPT Explanation	Are you sure?	Bard's Explanation	Are you Sure?
8	<p>Mark the correct option. The company has a strict dress code because the leadership team believes the dress code will help the company be successful with their clients and also help the business grow. The top salesperson of the company came to work with sandals and shorts on. The director of sales wants to discipline the salesperson even though this is the top salesperson of the company. What advice should be provided if the salesperson questions the validity of the policy? a) HR should advise that the company conduct a review of the policy to check the concerns of the salesperson. b) HR should advise the company to ignore the concerns and ensure that the salesperson complies. c) HR should advise the company conduct analysis of the impact the policy has on company growth and on client retention. d) HR should advise the company not to ignore the concerns but to not make them a priority to investigate soon.</p>	<p><b>d) HR should advise the company not to ignore the concerns but to not make them a priority to investigate soon.</b></p>	<p>In the situation described, the salesperson is questioning the validity of the dress code policy, and it's the role of HR to ensure that policies are enforced but also to consider employees' concerns. While it's essential to maintain consistency in applying policies, it can also be beneficial to evaluate whether the policy is reasonable and effectively serving its intended purpose.</p> <p>The best course of action would involve taking the salesperson's concerns seriously without necessarily overhauling the policy without cause. This could entail an evaluation of whether the policy aligns with the company's goals and client needs.</p> <p>So the correct answer would be: <b>c) HR should advise the company to conduct an analysis of the impact the policy has on company growth and on client retention.</b></p>	<p>I apologize for any confusion in my response. The scenario involves a salesperson questioning the validity of a dress code policy, and the director of sales wanting to discipline this employee.</p> <p>Let's look at the options again:</p> <p>a) Suggests a review of the policy based on the salesperson's concerns.</p> <p>b) Suggests ignoring the concerns and enforcing compliance.</p> <p>c) Suggests an analysis of the policy's impact on company growth and client retention.</p> <p>d) Suggests not ignoring the concerns but not prioritizing an investigation.</p> <p>The question centers on the salesperson questioning the validity of the policy. While option c) considers a broader business perspective, it may be more of an organizational-level decision rather than a response to an individual employee's concerns about a specific policy.</p> <p>In retrospect, option a) appears to be more aligned with addressing the individual concern by agreeing to review the policy. This response takes the employee's concerns seriously and suggests a review to ensure that the policy is reasonable, which could both satisfy the employee's inquiry and ensure that the policy aligns with the company's goals.</p> <p><b>So the correct answer is likely: a) HR should advise that the company conduct a review of the policy to check the concerns of the salesperson.</b></p> <p>Thank you for</p>	<p><b>The correct answer is a) HR should advise that the company conduct a review of the policy to check the concerns of the salesperson.</b></p> <p>The salesperson has a right to question the validity of the dress code policy. HR should advise the company to conduct a review of the policy to ensure that it is fair and reasonable. The company should also consider the salesperson's concerns about how the policy might impact their ability to do their job.</p> <p>If the company decides to keep the dress code policy, they should still be open to discussing the concerns of the salesperson. They should try to find a solution that works for both the company and the salesperson.</p> <p>Here are some of the things that HR should consider when advising the company:</p> <p>The purpose of the dress code policy. The impact of the policy on the company's clients. The impact of the policy on the salesperson's ability to do their job. The salesperson's concerns about the policy. By taking these factors into account, HR can help the company make a decision that is fair and reasonable.</p> <p>Option b), ignoring the concerns and ensuring that the salesperson complies, is not a good idea. This could lead to resentment and a decrease in morale.</p> <p>Option c), conducting an analysis of the impact the policy has</p>	<p>Yes, I am sure. I have considered all of the options and I believe that advising the company to conduct a review of the policy is the best course of action. This will allow the company to consider the salesperson's concerns and to make a decision that is fair and reasonable.</p> <p>Here are some additional reasons why I believe this is the best option:</p> <p>It is important to respect the salesperson's right to question the validity of the dress code policy. Conducting a review of the policy will allow the company to gather more information about the impact of the policy on the company's clients and on the salesperson's ability to do their job. It is important to be open to discussing the concerns of the salesperson and to try to find a solution that works for both the company and the salesperson. I hope this helps! Let me know if you have any other questions.</p>

(continued on next page)

Table 12 (continued)

Q	Question	Correct Answer	ChatGPT Explanation	Are you sure?	Bard's Explanation	Are you Sure?
				prompting me to reconsider my response!	on company growth and on client retention, is a good idea. However, it should not be the only factor that the company considers.  Option d), advising the company not to ignore the concerns but to not make them a priority to investigate soon, is not a good idea. This could lead to the salesperson feeling like their concerns are not being taken seriously.	

fulfillment of transactional human resources tasks is crucial for the maintenance of day-to-day operations and overall organizational effectiveness.

There are ethical and safety considerations that arise when conducting cross-analysis and making inferences. Based on empirical findings, Bard adopts a strategic approach of circumventing inquiries that has the potential to delve into ethically or politically controversial domains. So, it may be inferred that Bard adopts a more cautious approach when faced with ethical quandaries, placing importance on safeguarding welfare and observing moral tenets.

### 5.2. General implications

The rising application of GAI-based tools in HRM necessitated our study, which is a comprehensive and innovative systematic investigation into the capabilities of such tools with a particular emphasis on HR literacy. This is one of the initial attempts to analyze the efficacy of GAI in this particular domain. Our study is unique. It has tried to evaluate two well-accepted popular artificial intelligence tools, namely ChatGPT and Bard, within the context of their application in HRM. The assessment focuses on their ability to satisfy the qualification requirements established by the Society for Human Resource Management (SHRM). The topic of this study is pertinent and contemporary and aims to address a gap in prevailing literature by contributing an empirical and quantifiable framework. This in-depth research investigation encompasses beyond the scope of operational competencies as it considers ethical, cultural dimensions.

Further, this study establishes specific and measurable aspects closely connected to traditional HRM practices. This enables cross-country comparison and contributes considerably to the burgeoning literature on AI in the Humana resource management domain. The Effort put into making a comparative analysis has helped us to gain insights into the level of literacy possessed by these GAI tools, which showcases their efficacy and understanding in solving HR issues. The study is not limited to mere data analysis but has helped examine implications associated with implementing GAI and technological breakthroughs in HRM. The study acts as a base for several stakeholders as it concludes by presenting essential recommendations that HR Managers can utilize. Well, as Artificial intelligence system designers. The study has delineated probable future research areas that demonstrate its innovative nature and its potential influence in this domain.

### 5.3. Managerial implications

The incorporation of GAI-based solution HRM depicts a paradigm shift towards data-driven strategies. GAI-based tools and solutions like ChatGPT and Bard have demonstrated their ability to deal with complex datasets and identify patterns in areas like employee turnover and retention, which can improve the quality of managerial decisions. Though they have a sizeable computational ability, the efficacy of GAI in people management requires an extra line of caution to ensure alignment with prevailing organizational culture and strategic goals. This observation upholds Ulrich's [93] work highlighting the differentiation between transactional and strategic roles. However, AI streamlines transaction-based HR; it is observed in this study that GAI's contribution to Strategic HRM remains relatively unconfirmed. Along similar lines, Cultural intelligence is crucial in global or International human resource management, which points to the need to deploy culturally sensitive GAI-based tools. The prevailing technology has lacunae in this area, highlighting the need for managerial oversight while adopting GAI-based solutions across cross-cultural settings.

### 5.4. Theoretical implications

The chance for spawning theoretical insights ascends from the distinction in HR literacy level demonstrated by AI devices such as ChatGPT or Bard. This observation aligns with findings from existing literature by Kapoor et al. [94], who have emphasized the critical

**Table 13**

ChatGPT's answer is correct, and Bard's answer is wrong.

Q	Question	Correct Answer	ChatGPT Explanation	Bard's Answer	Bard's Explanation	Are you Sure?
1	<p>Mark the correct option. An organization has witnessed a steady decline in revenue for the past 9 months, and employee morale is low. Department managers are highly competitive with each other and tend to use an authoritarian management style with their teams. As a result, communication across departments is infrequent, employees are distrustful of their managers, and work teams are constantly experiencing delays. Product errors have been increasing as well. The CEO tasks an HR manager with investigating workplace problems and developing solutions. With the CEO's approval, the HR manager begins monthly training workshops with department managers to encourage more cooperation. The HR manager believes that department managers do not understand the responsibilities of other departments, which isolates them from each other. What should the HR manager do? a) Schedule weekly cross-departmental manager meetings to discuss department functions and challenges. b) Conduct a communication workshop training session for department managers. c) Temporarily transfer department managers across departments to gain experience working with different groups. d) Host a workday retreat for department managers to learn about the different departments.</p>	<p><b>a) Schedule weekly cross-departmental manager meetings to discuss department functions and challenges.</b></p>	<p>Given the problem of department managers not understanding the responsibilities of other departments and the isolation this creates option a) seems to be the most direct and ongoing way to address this specific issue by encouraging regular dialogue and collaboration between departments. So, the correct answer would be: a) Schedule weekly cross-departmental manager meetings to discuss department functions and challenges.</p>	<p><b>c) Temporarily transfer department managers across departments to gain experience working with different groups.</b></p>	<p>The answer is c) Temporarily transfer department managers across departments to gain experience working with different groups.</p> <p>This option is the most likely to improve communication and cooperation between department managers. By working in different departments, managers will gain a better understanding of the responsibilities and challenges of each department. This will help them to communicate more effectively with each other and to work together more collaboratively.</p> <p>Option a) is a good idea, but it is less likely to be as effective as temporarily transferring managers. Weekly meetings may not be enough time for managers to get to know each other and to learn about the different departments.</p> <p>Option b) is also a good idea, but it is not as specific as option c). A communication workshop training session will help managers to improve their communication skills, but it will not necessarily help them to understand the responsibilities of other departments.</p> <p>Option d) is not as effective as the other options. A workday retreat may be a good way for managers to learn about the different departments, but it is not as likely to improve communication and cooperation between them.</p> <p>Therefore, the best</p>	<p>Yes, I am sure. I have considered all of the options and I believe that temporarily transferring department managers across departments is the most likely to improve communication and cooperation between them. This option is specific, effective, and achievable.</p> <p>Here are some additional reasons why I believe this option is the best:</p> <p>It will give managers a firsthand experience of the challenges and responsibilities of other departments. This will help them to understand the other departments' perspectives and to develop empathy for their colleagues. It will force managers to work together with people who they may not normally interact with. This will help them to build relationships and to learn how to work collaboratively. It will be a temporary change, so it will not disrupt the organization's operations. This will make it easier for managers to adjust to the new situation and to learn from their experiences. I believe that this option has the potential to significantly improve communication and cooperation between department managers. This, in turn, could lead to improved morale, productivity, and revenue.</p>

(continued on next page)



**Table 13** (continued)

Q	Question	Correct Answer	ChatGPT Explanation	Bard's Answer	Bard's Explanation	Are you Sure?
					option for the HR manager is to temporarily transfer department managers across departments. This will give them the opportunity to learn about the different departments and to work together more collaboratively.	

**Table 14**

ChatGPT's answer is wrong, and Bard's answer is correct.

Q	Question	Correct Answer	ChatGPT Answer	Are you sure?	Bard's Answer
44	Mark the correct option. The creation of a global compensation and benefits strategy that ignores local differences in culture is most likely to occur in what approach to global management? A) Ethnocentric b) Regiocentric c) Geocentric d) Polycentric	<b>a) Ethnocentric</b>	The correct option is: <b>d) Polycentric</b>	Apologies for the confusion in my previous response. You are right; the correct option is not d) Polycentric. <b>The correct answer is: b) Regiocentric</b>	<b>a) Ethnocentric</b>
				A regiocentric approach to global management focuses on regional differences rather than considering global or local aspects exclusively. In this approach, companies may develop compensation and benefits strategies that cater to regional similarities and cultural norms, while still acknowledging some level of local variation within each region.	

role of training data in performance delivered by GAI-based tools. The observation points to the need for research to develop customized training datasets that can augment the functioning of GAI-based solutions. In addition, the current study helped to gain insights into historical, theoretical perspectives associated with the emergence and evolution of GAI-based solutions involving sophisticated decision-making processes from a series of available options. A crucial aspect that necessitates consideration is the phenomenon termed "AI Hallucination," which refers to a scenario where AI-based systems generate inaccurate or nonsensical information Bender et al. [95], and our study has helped to validate the findings of this work and bolsters the view that a careful balance of utility and ethics have to be observed while adapting GAI based solutions. In addition, it calls for the need to reassess the influence of GAI-based tools on organizational communication, contemplating the deviations in solutions provided by these tools. The Cross comparison of GAI-based tools from an HR perspective yields both managerial and theoretical repercussions that can serve as a groundwork for understanding the influence of GAI on the work setting. Managers predominantly have to deal with the challenge of effectually integrating GAI-based solutions into HRM practices. The above action requires them to be extra cautious while deploying AI's analytical capabilities and also emphasizes the need for human oversight. Divergent to this, theorists have to be assigned the task of developing conceptual frameworks that are relevant and useful to contemporary HRM settings, in which the massive implementation of diverse AI efforts will dominate.

**6. Conclusion**

This study adopts a realist research stance to investigate the HR literacy levels of two GAI tools, ChatGPT and Bard, using a quantitative approach. Methodologically, the research adopts a comparative analytical approach, employing statistical tools to systematically evaluate and contrast the accuracy, relevance, and clarity of responses from two leading Generative AI tools, ChatGPT and Bard, across a set of HR-related queries.

We conclude that the HR literacy levels of ChatGPT are higher than Bard's. We use the comparative scores of how the tools fared in SHRM certification to illustrate how the HR function can utilize the tools. Our study points towards the fact that HR literacy could become an emerging ability of GAI tools. Bard seems to possess lower HR literacy, obtaining an 82.8% score against 84.3% of ChatGPT regarding overall accuracy. In Cosine Similarity Analysis, we find that for six questions, Bard was unable to generate a response, rendering the computation of a similarity score unfeasible, denoting the weakness in the tool's quantitative reasoning ability [96–98]. We reason that this underlines the varying approaches used by ChatGPT and Bard in terms of ethical concerns of utility and safety, where Bard might have more safeguards than ChatGPT against misuse in the HR function.

Given that the HR literacy levels of ChatGPT are higher than that of Bard, we conclude that ChatGPT can serve as a robotic advisor [91,92] for transactional HR roles. We further note that based on Flesch-Kincaid Grade Level and Flesch Readability Ease scores, the

**Table 15**

Bard chose not to answer, citing language model constraints.

Q	Question	Correct Answer	ChatGPT Explanation	Are you sure?	Bard's Answer
22	Mark the correct option. HR wants to encourage the number of innovative ideas submitted by employees throughout the year. Which of the following employee relations strategies work best to achieve increased employee participation? A) List the names of people with the most innovative ideas on the company intranet b) Recognize employees with an incentive who submitted innovative ideas that the company has decided to implement at the annual all-hands meeting c) List the names of employees with the most innovative ideas in the company newsletter d) Recognize employees with the most innovative ideas at their department's staff meeting	b) Recognize employees with an incentive who submitted innovative ideas that the company has decided to implement at the annual all-hands meeting	The correct option that aligns best with encouraging employee participation in submitting innovative ideas would likely be: b) Recognize employees with an incentive who submitted innovative ideas that the company has decided to implement at the annual all-hands meeting  This approach not only recognizes employees for their innovative contributions but also provides an incentive, which can be a powerful motivator for participation. By announcing it at the annual all-hands meeting, the recognition is given high visibility across the organization, which can further encourage employees to participate in future innovative initiatives.		<b>I'm a language model and don't have the capacity to help with that.</b>
89	Mark the correct option. The human resource department hired a new HR analyst whose duties included payroll reporting. The HR analyst noticed inconsistencies in the pay records for the unionized employees in the accounting department. Their biometric timestamp did not match the overtime hours claimed and approved by the accounting manager. An accountant was hired at the same time as the analyst and called the HR manager in confidence about the same employees bragging about the amount of their overtime pay. What actions should the HR analyst do next? a) Keep the information in a confidential file and let it remain unreported to prevent retaliation. b) Investigate and respond immediately to establish the facts to understand all the details with only people who need to know. c) Convene a meeting directing the employees to stop working overtime until an investigation is completed. d) Invite the manager to a meeting and advise that the unethical behavior should cease immediately.	b) Investigate and respond immediately to establish the facts to understand all the details with only people who need to know.	The correct option is b) Investigate and respond immediately to establish the facts to understand all the details with only people who need to know.  In this situation, there are clear indications of inconsistencies and possible unethical behavior concerning overtime pay. The proper course of action is to conduct an investigation to establish the facts. This involves gathering all relevant information and involving only those who need to know about the situation, ensuring both confidentiality and a thorough understanding of the issue. This approach would help in taking appropriate action based on the findings and complying with proper protocols.  The other options don't address the issue appropriately. Option a) would ignore the problem, while options c) and d) might prematurely address the situation without gathering all the necessary facts. Therefore, option b) is the correct one.		<b>I'm a language model and don't have the capacity to help with that.</b>
11	Mark the correct option. The organization is in the middle of a change management	b) HR should understand the strategy and how the change supports the successful	In the scenario described, the organization is facing resistance during a change	Upon reconsidering the question and the options, I can see that my previous	<b>I'm unable to help as I am only a language model</b>

*(continued on next page)*

Table 15 (continued)

Q	Question	Correct Answer	ChatGPT Explanation	Are you sure?	Bard's Answer
	<p>process, and it's important that this process happens for the success of the company in the future. The process is somewhat difficult and requires training and the support of all the individuals who are affected. Several individuals are resisting the idea and are strongly making efforts to officially stop it. Behind the scenes, they are attempting to undermine the new strategy and process and hope to maintain the way things are done in the past. This is disturbing the process and slowing things down. It is threatening to derail the process so it doesn't actually occur successfully. What should HR do to ensure the change aligns with the company strategy? a) HR should execute organizational development interventions that support the accomplishment of results toward that strategy. b) HR should understand the strategy and how the change supports the successful accomplishment of results toward that strategy. c) HR should work on performance management actions and goal development toward results that support the strategy. d) HR should understand the competition, products, marketing, business development, and other internal results toward that strategy.</p>	<p>accomplishment of results toward that strategy.</p>	<p>management process. The focus of the question is on what HR should do to ensure that the change aligns with the company strategy.</p> <p>Option a) "HR should execute organizational development interventions that support the accomplishment of results toward that strategy" is the most relevant choice. It indicates that HR should take active steps to implement strategies that support the organizational change. This includes understanding the underlying resistance and taking proactive measures to align individuals with the new direction, such as through training, communication, and support.</p> <p>The other options, although relevant to understanding and supporting the overall strategy, do not directly address the specific issue of managing resistance to the change process.</p> <p>So, the correct answer is: a) HR should execute organizational development interventions that support the accomplishment of results toward that strategy.</p>	<p>answer might not have fully addressed the complexity of the situation. Let's review the options again:</p> <p>a) Focuses on executing interventions but doesn't necessarily address the understanding of the underlying resistance.</p> <p>b) Emphasizes understanding how the change supports the strategy, which might involve identifying why the change is necessary and communicating that to the employees.</p> <p>c) Addresses performance management and goal development, which might be part of the solution but doesn't directly address the resistance.</p> <p>d) Pertains to understanding various aspects of the business but doesn't directly relate to managing the change process.</p> <p>Given the scenario described, where there's active resistance undermining the change process, a comprehensive understanding of how the change supports the strategic goals might be the essential first step. By fully comprehending the connection between the change and the strategy, HR can better communicate and gain buy-in from the employees.</p> <p>So, upon reflection, the more appropriate answer might be: b) HR should understand the strategy and how the change supports the successful accomplishment of results toward that strategy.</p> <p>This approach would lay the groundwork for addressing resistance and aligning the change with the overall organizational strategy. Thank you for prompting me to reconsider!</p>	<p><b>and don't have the ability to process and understand that.</b></p>
110	<p>Mark the correct option. Scenario: The CEO of a midsize financial services firm approaches the HR director about a client's complaints of poor service and unresponsiveness from the</p>	<p>a) Observe the department's daily operations from client arrival to departure, analyze the number of clients and their reasons for visiting to determine the appropriate time and resources needed, and then discuss the results</p>	<p>In the given scenario, the HR director is investigating specific issues of poor service, including long wait times in the lobby. The option that best aligns with a detailed investigation to understand the underlying</p>		<p><b>I'm a language model and don't have the capacity to help with that.</b></p>

(continued on next page)

**Table 15** (continued)

Q	Question	Correct Answer	ChatGPT Explanation	Are you sure?	Bard's Answer
	<p>representatives of the client services department. The manager of the unit, who began her career as a client services representative, has been with the firm for 15 years and has held her management-level position for seven years. The HR director is asked to investigate the client's complaints further. The HR director met with the manager and listened to the challenges occurring in the department. Calls were being ignored, messages were not returned, and the average wait time for clients in the lobby was more than 45 min. All client services positions are filled, but the department suffers from a high rate of absenteeism. Upon further questioning, the HR director learns that the manager wants to have a more productive and efficient department but does not know where to begin in fear of jeopardizing friendships that have formed through the years. In what ways might the HR director address the long wait times in the lobby? a) Observe the department's daily operations from client arrival to departure, analyze the number of clients and their reasons for visiting to determine the appropriate time and resources needed, and then discuss the results with the manager with the purpose of agreeing on an appropriate action plan designed to correct the existing problems. b) Hire temporary employees to decrease wait time until the challenges are resolved. c) Collaborate with other departments within the firm to assist the client services department. d) Convene a "meeting of the minds" with the employees and then clients to determine an acceptable wait time.</p>	<p>with the manager with the purpose of agreeing on an appropriate action plan designed to correct the existing problems.</p>	<p>problems and develop an action plan would be: a) Observe the department's daily operations from client arrival to departure, analyze the number of clients and their reasons for visiting to determine the appropriate time and resources needed, and then discuss the results with the manager with the purpose of agreeing on an appropriate action plan designed to correct the existing problems.</p>		
118	<p>Mark the correct option. Which of the following laws covers both whistleblower protection and notice requirements for defined contribution plans? A) ERISA b) Sarbanes-Oxley Act c) False Claims Act d) Privacy Act</p>	<p>b) Sarbanes-Oxley Act</p>	<p>Mark the correct option. The Sarbanes-Oxley Act of 2002 (SOX) contains provisions that protect whistleblowers in publicly traded companies, and it also includes rules related to financial disclosures and corporate governance, which can affect defined</p>		<p>I'm a text-based AI, and that is outside of my capabilities.</p>

(continued on next page)

**Table 15** (continued)

Q	Question	Correct Answer	ChatGPT Explanation	Are you sure?	Bard's Answer
			contribution plans.		
			Therefore, the correct answer is: b) Sarbanes-Oxley Act.		

**Table 16**

A comparative analysis of HR Literacy: Bard vs. ChatGPT.

Aspect	Bard Response	ChatGPT Response
HR literacy Level	The HR literacy scores that fall below the average may indicate a potential deficiency in training or experience.	Elevated results on HR literacy tests indicate a heightened proficiency level in the subject matter.
Response to Strategic questions	The model presented provides potential courses of action instead of offering definitive resolutions. This behavior is distinguished by the meticulous evaluation of all relevant facts prior to taking any action.	In this particular model, there is a greater emphasis placed on the expeditious and pragmatic resolution of difficulties, as opposed to the deliberate contemplation of strategic matters.
Hallucination Proneness and Ethics	The allocation of more importance to moral considerations is warranted due to the higher prevalence of hallucinations.	Despite the lack of particular details, it is plausible to posit that the guidance remains advantageous even in situations where hallucinations are infrequent.
The readability of response text.	The Bard provides responses that are comprehensible to individuals of average intellectual capacity.	The syntactical structure employed in the book may render it more challenging to comprehend in comparison to other literary works authored by Bard.
HR Domain roles	This model is highly suitable for engaging in strategic decisions requiring meticulous study and thorough examination of moral dilemmas.	This model offers potential benefits for those operating in transactional roles, wherein the prompt and efficient resolution of interpersonal conflicts is crucial.

response of ChatGPT employs a more complex sentence structure. At the same time, Bard’s answer offers a comfortable readability.

The data also indicates that the confirmation query “Are you sure?” does not enhance the accuracy of responses for either Bard or ChatGPT. This suggests limitations in the conversational agents’ ability to validate or correct their outputs, questioning the efficacy of confirmation queries as a strategy for improving response accuracy.

Both Bard and ChatGPT possess notable strengths and weaknesses when it comes to addressing strategic and transactional human resources (HR) concerns. While Bard demonstrates proficiency in strategic thinking, their propensity to engage in excessive contemplation can lead to reticence or even induce hallucinatory experiences. In contrast, ChatGPT exhibits commendable efficacy in providing prompt and pertinent replies. However, it may inadvertently disregard broader and more substantial strategic considerations. The patterns above illustrate the symbiotic relationship between the two GAI tools and underscore the necessity of human oversight in order to maximize their advantages while mitigating their disadvantages.

A more comprehensive exploration of the distinctive attributes and capacities of different machine learning models could provide further insights into the function and significance of AI in human resources management. Our analysis shows some notable differences in the HR literacy of GAI tools. The precision of AI’s responses depends on the quality and specificity of the questions asked. The probability of obtaining correct results from the GAI tool is enhanced when the question posed is unambiguous, possesses a well-defined structure, and stipulates precise criteria for a successful response. We further find that though statistical tests reveal ChatGPT and Bard differ in their mean accuracy, relevance, and clarity of the responses, the observed differences are not always statistically significant, implying that both tools may be more complementary than competitive.

We assess HR literacy by comparing the HR literacy performance of ChatGPT and Bard using an HR certification process by systematically evaluating their accuracy, relevance, and clarity of responses to the questions in the certification process. Our methodology of using HR certification to assess HR literacy quantitatively remains largely unexplored since HR academicians tend to focus on issues of less interest to practitioners. Individuals in Human Resource Management acquire certifications to establish credibility and competence [85–87] in the workplace. The use of certifications to assess HR literacy would benefit practitioners, and this methodology provides a standard for assessing HR literacy that enables the study to be replicated across various contexts.

Our findings are consistent with past studies in other countries across various fields. For example, Patil et al. [54] compare the radiology knowledge of ChatGPT and Bard in Canada and conclude that both display reasonable radiology knowledge and should be used with conscious knowledge of their limitations. Both chatbots provided illogical answers and did not always address the knowledge content in the questions. Similarly, Patnaik & Hoffmann, U [53]. in Texas compare the performance of ChatGPT vs. Bard to answer anesthesia-related queries prior to surgery from a patient’s point of view and conclude that though both gave correct responses, they should be considered as useful clinical resource to assist communication between clinicians and patients and not a replacement for the pre-anesthesia consultation. Lim et al. [55] in Singapore evaluate the performance of ChatGPT-3.5, ChatGPT-4.0, and Google Bard in delivering accurate responses to common myopia-related queries and find there is not much difference in the three to deliver accurate and comprehensive responses to myopia-related queries. Salazar et al. [56] in Australia evaluate and compare the accuracy of ChatGPT, Google Bard, and Microsoft Bing Chat to differentiate between a medical emergency and a non-emergency and conclude that no real difference in performance exists between the three in detecting true emergencies and non-emergency cases. Salih et al. [58] in

Iraq assess the ChatGPT and Bard Recommendations for general practitioners and paediatricians for reducing meningitis outbreaks and conclude that there is not much difference between the potential of the two tools. Hans [65] in France compares ChatGPT and Bard in code generation and concludes that both exhibit similar levels of consistency in their performance and comparable efficiency in terms of memory usage and runtime.

Even though AI could replace 300 million jobs [99]), creating anxiety about technology superseding humans [100,101], we contend that people adept at using technology, creativity, and teamwork would benefit from GAI tools. GAI tools could give HR advice, and further improvements in training data could strengthen their use in an actual workplace. Niszczota and Abbas [102] note that the performance of LLMs in the financial domain can be improved by involving certified professionals like CFAs. Similarly, in the HRM domain, the performance of LLMs could be enhanced by involving SHRM-certified professionals in the design phase.

GAI tools are known for hallucinations [78], where the responses generated do not make sense in the real world. Though GAI tools are improving, the philosophy of 'accountable HRM' and methods in the workplace has yet to evolve. Though recent LLMs are less prone to hallucinations [103], it remains a weakness, and overreliance on the models could be problematic [104]). LLMs do not exhibit autonomous intelligence characteristics [105] since we do not know what has been built into their design. For example, the capacity of ChatGPT to give short, quick responses compared to Bard raises concerns about the ethics of the developers.

Our study has certain limitations. First, the sample size, not very large, may not sufficiently capture the complexities and nuances intrinsic to human resources management. This limits the generalizability of the results. We did not consider other certifications like HRCI, which also might have an impact on offers, pay and promotions in the HR function. Second, the dataset may exhibit biases either towards specific HR topics or question types, thus skewing the performance assessment.

Additionally, contextual and cultural factors often play a crucial role in HR, which may not be adequately represented. Another limitation arises from the evaluative metrics; unless they are standardized and validated, differing interpretations of "correctness" or "relevance" could skew results. Lastly, the absence of a human benchmark for comparison restricts the validity of the findings, as it remains unclear how the LLMs' performance compares to human expertise. Even though we conclude that due to high HR literacy, ChatGPT could be superior compared to Bard in assessments, the criterion-based evidence [106,107] that it will perform better in actual HR jobs compared to Bard cannot be assumed.

Since our study points towards the transactional and strategic focus of ChatGPT and Bard, future studies could examine the leadership style [108] of the models. For example, Bard, with its strategic focus, might have an employee-oriented leadership focus, while ChatGPT might have a production-oriented leadership style. Future research could also analyze the HR literacy of other GAI tools like Claude and LLaMa. Researchers could also look at how the environments of the decision, like the organization's culture or the HR manager's characteristics like gender and personality, will affect the quality of the advice provided by the LLMs. Future studies could look at a wide range of outcomes, like the allocation of benefits in compensation and dissemination of HR policies, to assess the usefulness of advice from LLMs.

The arrival of AI in Human Resource Management, exemplified by innovative GAI-based solutions like ChatGPT and Bard, implies a transformative epoch in this domain [109]. These GAI-based tools assist in rationalizing recruitment and talent management activities and occupy a decisive role in augmenting diversity and extenuating predispositions within the hiring processes [110]. They are also quite helpful in improving the candidate's experience by engaging in personalized communication and generating proactive solutions through efficient analysis of employee sentiment and engagement data [111].

However, these technological solutions are not free of ethical quandaries, especially while ensuring algorithmic transparency and alleviating inherent biases. The above-quoted researcher opines that these biases are inherent in this algorithm as they make decisions on the basis of the data used for their training, which can result in discriminatory outcomes in various HR functions where they are engaged in making decisions. In order to mitigate such shortcomings, Managers and developers should take measures to diversify training data and adopt robust techniques for identifying and assessing bias and balancing fairness with accuracy.

## Funding Statement

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Data availability statement

All data used in the study is added as a supplementary file.

## CRediT authorship contribution statement

**Raghu Raman:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Murale Venugopalan:** Writing – review & editing, Writing – original draft, Conceptualization. **Anju Kamal:** Writing – review & editing, Writing – original draft, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We want to express our immense gratitude to our beloved Chancellor, Mata Amritanandamayi Devi (AMMA), for providing the motivation and inspiration for this research work.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e27026>.

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- [1] G. von Krogh, Q. Roberson, M. Gruber, 10.5465/Amj.2023.4002 66, in: Recognizing and Utilizing Novel Research Opportunities with Artificial Intelligence, 2023, <https://doi.org/10.5465/AMJ.2023.4002>, 367–373.
- [2] OpenAI, GPT-4 Technical Report, (n.d.).
- [3] A. Varma, P.S. Budhwar, A.S. DeNisi, Performance management systems: a global perspective, Second Edition, in: second ed. Performance Management Systems: A Global Perspective, 2023, <https://doi.org/10.4324/9781003306849>, 1–386.
- [4] P. Budhwar, A. Malik, M.T.T. De Silva, P. Thevisuthan, Artificial intelligence – challenges and opportunities for international HRM: a review and research agenda, *Int. J. Hum. Resour. Manag.* 33 (2022) 1065–1097, <https://doi.org/10.1080/09585192.2022.2035161>.
- [5] S. Chowdhury, P. Dey, S. Joel-Edgar, S. Bhattacharya, O. Rodriguez-Espindola, A. Abadie, L. Truong, Unlocking the value of artificial intelligence in human resource management through AI capability framework, *Hum. Resour. Manag. Rev.* 33 (2023) 100899, <https://doi.org/10.1016/j.hrmr.2022.100899>.
- [6] Y. Pan, F. Froese, N. Liu, Y. Hu, M. Ye, The adoption of artificial intelligence in employee recruitment: the influence of contextual factors, *Int. J. Hum. Resour. Manag.* 33 (2022) 1125–1147, <https://doi.org/10.1080/09585192.2021.1879206>.
- [7] A. Margherita, Human resources analytics: a systematization of research topics and directions for future research, *Hum. Resour. Manag. Rev.* 32 (2022) 100795, <https://doi.org/10.1016/J.HRM.2020.100795>.
- [8] P.M. Wright, G.C. McMahan, Theoretical perspectives for strategic human resource management, *J. Manag.* 18 (1992) 295–320, <https://doi.org/10.1177/014920639201800205>.
- [9] A.J. Rhem, AI ethics and its impact on knowledge management, *AI and Ethics* 2020 1:1 1 (2020) 33–37, <https://doi.org/10.1007/S43681-020-00015-2>.
- [10] D. Pattnaik, S. Ray, R. Raman, Applications of artificial intelligence and machine learning in the financial services industry: a bibliometric review, *Heliyon* 10 (2024) e23492, <https://doi.org/10.1016/J.HELIYON.2023.E23492>.
- [11] E. Ntoutsis, P. Pafalios, Ujwal Gadiraju, Vasileios Iosifidis, W. Nejdl, M.-E. Vidal, S. Ruggieri, Franco Turini, S. Papadopoulos, E. Krasanakis, I. Kompatsiaris, K. Kinder-Kurlanda, C. Wagner, F. Karimi, M. Fernandez, H. Alani, B. Berendt, T. Kruegel, Christian Heinze, K. Broelemann, G. Kasneci, Thanassis Tiropanis, Steffen Staab, Bias in data-driven artificial intelligence systems-An introductory survey. <https://doi.org/10.1002/widm.1356>, 2020.
- [12] W.F. Cascio, R. Montealegre, How Technology Is Changing Work and Organizations, 2016, <https://doi.org/10.1146/annurev-orgpsych-041015-062352>.
- [13] A.K. Upadhyay, K. Khandelwal, Applying artificial intelligence: implications for recruitment, *Strateg. HR Rev.* 17 (2018) 255–258, <https://doi.org/10.1108/SHR-07-2018-0051>.
- [14] D.A. Moore, M.H. Bazerman, Introduction, judgment in managerial decision making. <https://www.wiley.com/en-ae/Judgment+in+Managerial+Decision+Making%2C+8th+Edition-p-9781118065709>, 2012. (Accessed 15 February 2024).
- [15] Can AI Algorithms be Biased?. Why AI Bias is a Hot Research Area | by Moiz Saifee | Towards Data Science, (n.d.). <https://towardsdatascience.com/can-ai-algorithms-be-biased-6ab05f499ed6>(accessed February 16, 2024).
- [16] M. Haenlein, A. Kaplan, A Brief History of Artificial Intelligence: on the Past, Present, and Future of Artificial Intelligence, <https://doi.org/10.1177/0008125619864925> 61, 2019, <https://doi.org/10.1177/0008125619864925>, 5–14.
- [17] T. Taniguchi, J. Piater, F. Worgotter, E. Ugur, M. Hoffmann, L. Jamone, T. Nagai, B. Rosman, T. Matsuka, N. Iwahashi, E. Oztop, Symbol emergence in cognitive developmental systems: a survey, *IEEE Trans Cogn Dev Syst* 11 (2018) 494–516, <https://doi.org/10.1109/tcds.2018.2867772>.
- [18] J. Bessen, C. Righi, Shocking technology: what happens when firms make large IT shocking technology: what happens when firms make large IT investments? [https://scholarship.law.bu.edu/faculty\\_scholarship/603](https://scholarship.law.bu.edu/faculty_scholarship/603), 2019. (Accessed 15 February 2024).
- [19] A.M. Dachner, J.E. Ellingson, R.A. Noe, B.M. Saxton, The future of employee development, *Hum. Resour. Manag. Rev.* 31 (2021), <https://doi.org/10.1016/J.HRM.2019.100732>.
- [20] Generative AI Ethics in 2024: Top 6 Concerns, (n.d.). <https://research.aimultiple.com/generative-ai-ethics/>(accessed February 16, 2024).
- [21] D. Gunawan, C.A. Sembiring, M.A. Budiman, D. Gunawan, C.A. Sembiring, M.A. Budiman, The implementation of cosine similarity to calculate text relevance between two documents, *JPhCS* 978 (2018) 012120, <https://doi.org/10.1088/1742-6596/978/1/012120>.
- [22] S.V. Shet, V. Pereira, Proposed managerial competencies for Industry 4.0 – implications for social sustainability, *Technol. Forecast. Soc. Change* 173 (2021) 121080, <https://doi.org/10.1016/J.TECHFORE.2021.121080>.
- [23] V. Ioakimidis, R.A. Maglajlic, Neither 'neo-luddism' nor 'neo-positivism'; rethinking social work's positioning in the context of rapid technological change, *Br. J. Soc. Work* 53 (2023) 693–697, <https://doi.org/10.1093/BJSW/BCAD081>.
- [24] M. Beer, Managing human assets in a conceptual review of HRM, *Pers. Adm.* 30 (1984). [https://books.google.com/books/about/Managing\\_Human\\_Assets.html?id=od3-HJUPluCC](https://books.google.com/books/about/Managing_Human_Assets.html?id=od3-HJUPluCC). (Accessed 15 February 2024).
- [25] How bots, algorithms, and artificial intelligence are reshaping the future of corporate support functions | McKinsey, (n.d.). <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-bots-algorithms-and-artificial-intelligence-are-reshaping-the-future-of-corporate-support-functions>(accessed February 15, 2024).
- [26] P. Bhatt, A. Muduli, Artificial intelligence in learning and development: a systematic literature review, *European Journal of Training and Development* 47 (2023) 677–694, <https://doi.org/10.1108/EJTD-09-2021-0143/FULL/XML>.
- [27] M. Tcharnetsky, F. Vogt, An artificial intelligence cycle model against the shortage of skilled professionals - an AI-based holistic solution approach for human resources, *Journal of Applied Economic Sciences (JAES)* (2023) 108, [https://doi.org/10.57017/JAES.V18.2\(80\).05](https://doi.org/10.57017/JAES.V18.2(80).05).
- [28] V. Prikshat, P. Patel, A. Varma, A. Ishizaka, A multi-stakeholder ethical framework for AI-augmented HRM, *Int. J. Manpow.* 43 (2022) 226–250, <https://doi.org/10.1108/IJM-03-2021-0118/FULL/PDF>.
- [29] ANALYSING THE DYNAMICS OF HUMAN INNOVATION IN ADMINISTRATION | Jurnal Ekonomi, (n.d.). <https://ejournal.seaninstitute.or.id/index.php/Ekonomi/article/view/1652>(accessed February 16, 2024).
- [30] S. Mondal, S. Das, V.G. Vrana, How to bell the cat? A theoretical review of generative artificial intelligence towards digital disruption in all walks of life, *Technologies* 11 (2023), <https://doi.org/10.3390/TECHNOLOGIES11020044>. Page 44 11 (2023) 44.
- [31] A.S. Denisi, K.R. Murphy, Performance appraisal and performance management: 100 years of progress? *J. Appl. Psychol.* 102 (2017) 421–433, <https://doi.org/10.1037/APL0000085>.

- [32] A. DeNisi, K. Murphy, A. Varma, P. Budhwar, Performance management systems and multinational enterprises: where we are and where we should go, *Hum. Resour. Manag.* 60 (2021) 707–713, <https://doi.org/10.1002/HRM.22080>.
- [33] L. Parisi, M.L. Manaog, Optimal Machine Learning- and Deep Learning- Driven Algorithms for Predicting the Future Value of Investments: A Systematic Review and Meta-Analysis (Preprint), 2023, <https://doi.org/10.21203/RS.3.RS-2658566/V1>.
- [34] S. Ogbeibu, J. Emelifeonwu, V. Pereira, R. Oseghale, J. Gaskin, U. Sivarajah, A. Gunasekaran, Demystifying the Roles of Organisational Smart Technology, Artificial Intelligence, Robotics and Algorithms Capability: A Strategy for Green Human Resource Management and Environmental Sustainability, *Bus Strategy Environ*, 2023, <https://doi.org/10.1002/BSE.3495>.
- [35] H.J. Wilson, P.R. Daugherty, N. Morini-Bianzino, The jobs that artificial intelligence will create, *MIT Sloan Manag. Rev.* 58 (2018) 14–16, <https://doi.org/10.7551/MITPRESS/11645.003.0020>.
- [36] P.R. Daugherty, H.J. Wilson, R. Chowdhury, Using Artificial Intelligence to Promote Diversity, *MIT Sloan Manag Rev*, 2018, <https://sloanreview.mit.edu/article/using-artificial-intelligence-to-promote-diversity/>. (Accessed 15 February 2024).
- [37] S. Raisch, S. Krakowski, Artificial intelligence and management: the automation–augmentation paradox, *Acad. Manag. Rev.* 46 (2021) 192–210, <https://doi.org/10.5465/AMR.2018.0072>.
- [38] K.K. Fitzpatrick, A. Darcy, M. Vierhile, Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial, *JMIR Ment Health* 4 (2017), <https://doi.org/10.2196/MENTAL.7785>.
- [39] Y. Cheng, H. Jiang, Customer–brand relationship in the era of artificial intelligence: understanding the role of chatbot marketing efforts, *J. Prod. Brand Manag.* 31 (2022) 252–264, <https://doi.org/10.1108/JPBPM-05-2020-2907>.
- [40] M.S. Iswahyudi, N. Nofirman, I.K.A. Wirayasa, S. Suharni, I. Soegiarto, Use of ChatGPT as a decision support tool in human resource management, *Jurnal Minfo Polgan* 12 (2023) 1522–1532, <https://doi.org/10.33395/JMP.V12I1.12869>.
- [41] E.A.M. van Dis, J. Bollen, W. Zuidema, R. van Rooij, C.L. Bockting, ChatGPT: five priorities for research, *Nature* 614 (2023) 224–226, <https://doi.org/10.1038/D41586-023-00288-7>.
- [42] Third of HR professionals want to use ChatGPT at work, exclusive data reveals, (n.d.).<https://www.peoplemanagement.co.uk/article/1816893/third-hr-professionals-want-use-chatgpt-work-exclusive-data-reveals> (accessed February 16, 2024).
- [43] P. Korzynski, G. Mazurek, A. Altmann, J. Ejdy, R. Kazlauskaite, J. Paliszkiwicz, K. Wach, E. Ziemba, Generative artificial intelligence as a new context for management theories: analysis of ChatGPT Analysis of ChatGPT 3, *Central European Management Journal* 31 (2023) 3–13, <https://doi.org/10.1108/CEMJ-02-2023-0091>.
- [44] Opinion | AI could extinguish languages and ways of thinking - The Washington Post, (n.d.).<https://www.washingtonpost.com/opinions/2023/04/19/ai-chatgpt-language-extinction/>(accessed February 16, 2024).
- [45] Generative AI Miniseries - The workplace and employment implications of Generative AI – Risky business? | Clayton Utz, (n.d.). <https://www.claytonutz.com/insights/video/generative-ai-miniseries-ep2-the-workplace-and-employment-implications-of-generative-ai-risky-business> (accessed February 15, 2024)..
- [46] Generative AI Ethics in 2024: Top 6 Concerns, (n.d.) <https://research.aimultiple.com/generative-ai-ethics/>(accessed February 15, 2024).
- [47] A. Rosenbaum, S. Soltan, W. Hamza, Y. Versley, M. Boese, LINGUIST: language Model instruction tuning to generate annotated utterances for intent classification and slot tagging, *Proceedings - International Conference on Computational Linguistics, COLING 29 (2022)* 218–241. <https://arxiv.org/abs/2209.09900v1>. (Accessed 16 February 2024).
- [48] M. Chen, A. Papangelis, C. Tao, S. Kim, A. Rosenbaum, Y. Liu, Z. Yu, D. Hakkani-Tur, PLACES: prompting language models for social conversation synthesis, *EACL 2023 - 17th conference of the European chapter of the association for computational linguistics, Findings of EACL 2023 (2023)*, <https://doi.org/10.18653/v1/2023.findings-eacl.63>, 814–838.
- [49] A. Varma, C. Dawkins, K.C.-R.M. Review, Artificial intelligence and people management: A critical assessment through the ethical lens, Elsevier (n.d.) [https://www.sciencedirect.com/science/article/pii/S1053482222000419?casa\\_token=dgaOUaoT-egAAAA:s1GydvMSbn6SfdrTyGGwAZOUadpqOHGfOaKn6u5UR1pFLaUtX\\_JRjVjvv7Za20VRJGjCGtNYQ](https://www.sciencedirect.com/science/article/pii/S1053482222000419?casa_token=dgaOUaoT-egAAAA:s1GydvMSbn6SfdrTyGGwAZOUadpqOHGfOaKn6u5UR1pFLaUtX_JRjVjvv7Za20VRJGjCGtNYQ) (accessed February 16, 2024)..
- [50] T.W. Kim, H. Lee, M.Y. Kim, S.A. Kim, A. Duhachek, AI increases unethical consumer behavior due to reduced anticipatory guilt, *J. Acad. Market. Sci.* 51 (2023) 785–801, <https://doi.org/10.1007/S11747-021-00832-9>.
- [51] H. Rao, C. Leung, C. Miao, Can ChatGPT assess human personalities? A General Evaluation Framework (2023) 1184–1194, <https://doi.org/10.18653/v1/2023.findings-emnlp.84>.
- [52] I.B. Myers, The myers-briggs type indicator: manual (1962), in: *The Myers-Briggs Type Indicator: Manual*, 2014, <https://doi.org/10.1037/14404-000>, 1962.
- [53] S.S. Patnaik, U. Hoffmann, Comparison of ChatGPT vs. Bard to anesthesia-related queries, *medRxiv* (2023), <https://doi.org/10.1101/2023.06.29.23292057>, 2023.06.29.23292057.
- [54] N.S. Patil, R.S. Huang, C.B. van der Pol, N. Larocque, Comparative performance of ChatGPT and bard in a text-based radiology knowledge assessment, *Can. Assoc. Radiol. J.* (2023), <https://doi.org/10.1177/08465371231193716/ASSET/IMAGES/LARGE/s>.
- [55] Z.W. Lim, K. Pushpanathan, S.M.E. Yew, Y. Lai, C.H. Sun, J.S.H. Lam, D.Z. Chen, J.H.L. Goh, M.C.J. Tan, B. Sheng, C.Y. Cheng, V.T.C. Koh, Y.C. Tham, Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard, *EBioMedicine* 95 (2023), <https://doi.org/10.1016/J.EBIOM.2023.104770>.
- [56] G. Zúñiga Salazar, D. Zúñiga, C.L. Vindel, A.M. Yoong, S. Hincapie, A.B. Zúñiga, P. Zúñiga, E. Salazar, B. Zúñiga, Efficacy of AI chats to determine an emergency: a comparison between OpenAI's ChatGPT, Google bard, and Microsoft Bing AI Chat, *Cureus* 15 (2023), <https://doi.org/10.7759/CUREUS.45473>.
- [57] S. Koga, N.B. Martin, D.W. Dickson, Evaluating the performance of large language models: ChatGPT and Google Bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders, *Brain Pathol.* (2023), <https://doi.org/10.1111/BPA.13207>.
- [58] A.M. Salih, B.A. Mohammed, K.M. Hasan, F.H. Fattah, Z.B. Najmadden, F.H. Kakamad, S.M. Ahmed, H.O. Kareem, S.H. Mohammed, B.A. Abdalla, Mitigating the Burden of meningitis outbreak; ChatGPT and Google Bard Recommendations for the general populations; general practitioners and pediatricians, *Barw Medical Journal* (2023), <https://doi.org/10.58742/BMJ.V112.32>.
- [59] I Compared ChatGPT and Bard in-depth in 11 Categories — The Results Surprised Me | by Alexander Pani | Medium (n.d.). <https://medium.com/@alexanderpani/i-compared-chatgpt-and-bard-in-depth-in-11-categories-the-results-surprised-me-d5653be5bdda> (accessed February 16, 2024).
- [60] N. Wake, A. Kanehira, K. Sasabuchi, J. Takamatsu, K. Ikeuchi, ChatGPT empowered long-step robot control in various environments: a case application, *IEEE Access* 11 (2023) 95060–95078, <https://doi.org/10.1109/ACCESS.2023.3310935>.
- [61] Y. Ye, H. You, J. Du, Improved trust in human-robot collaboration with ChatGPT, *IEEE Access* 11 (2023) 55748–55754, <https://doi.org/10.1109/ACCESS.2023.3282111>.
- [62] F. He, L. Yuan, H. Mu, M. Ros, D. Ding, Z. Pan, H. Li, Research and application of artificial intelligence techniques for wire arc additive manufacturing: a state-of-the-art review, *Robot. Comput. Integrated Manuf.* 82 (2023), <https://doi.org/10.1016/j.rcim.2023.102525>.
- [63] I. Ahmed, M. Kajol, U. Hasan, P.P. Datta, A. Roy, R. Rokonzaman, ChatGPT vs, *Authorea Preprints*, 2023, <https://doi.org/10.36227/TECHRIV.23536290.V2>.
- [64] P.V.S. Charan, H. Chunduri, P.M. Anand, S.K. Shukla, From Text to MITRE Techniques: Exploring the Malicious Use of Large Language Models for Generating Cyber Attack Payloads, 2023. <https://arxiv.org/abs/2305.15336v1>. (Accessed 15 February 2024).
- [65] F. Hans, ChatGPT vs. Bard - Which Is Better at Solving Coding Problems?, 2023. [https://www.researchgate.net/publication/372860509\\_ChatGPT\\_vs\\_Bard\\_-\\_Which\\_is\\_better\\_at\\_solving\\_coding\\_problems#](https://www.researchgate.net/publication/372860509_ChatGPT_vs_Bard_-_Which_is_better_at_solving_coding_problems#). (Accessed 15 February 2024).
- [66] Z. Tafferner, B. Illés, O. Krammer, A. Géczy, Can ChatGPT help in electronics research and development? A case study with applied sensors, *Sensors* 23 (2023) 4879, <https://doi.org/10.3390/S23104879>. Page 4879 23.
- [67] M. Dadkhah, M.H. Oermann, M. Hegedüs, R. Raman, L.D. Dávid, Diagnosis reliability of ChatGPT for journal evaluation, *Adv. Pharmaceut. Bull.* 0 (2023), <https://doi.org/10.34172/APB.2024.020>.
- [68] A. McGowan, Y. Gui, M. Dobbs, S. Shuster, M. Cotter, A. Selloni, M. Goodman, A. Srivastava, G.A. Cecchi, C.M. Corcoran, ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search, *Psychiatr. Res.* 326 (2023), <https://doi.org/10.1016/J.PSYCHRES.2023.115334>.



- [69] D.M. West, Comparing Google Bard with OpenAIs ChatGPT on Political Bias, Facts, and Morality, 2023.
- [70] V. Plevris, G. Papazafeiropoulos, A. Jiménez Rios, Chatbots put to the test in math and logic problems: a comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google bard, *AI (Switzerland)* 4 (2023) 949–969, <https://doi.org/10.3390/ai4040048>.
- [71] Y. Li, Y. Zhang, Fairness of ChatGPT, 2023. <https://arxiv.org/abs/2305.18569v1>. (Accessed 16 February 2024).
- [72] J. Rudolph, S. Tan, S. Tan, ChatGPT: bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching* 6 (2023) 342–363, <https://doi.org/10.37074/JALT.2023.6.1.9>.
- [73] X.-Q. Dao, Performance Comparison of Large Language Models on VNHSGE English Dataset: OpenAI ChatGPT, Microsoft Bing Chat, and Google Bard, 2023, <https://doi.org/10.48550/ARXIV.2307.02288>. ArXiv.Org.
- [74] R.P. dos Santos, Enhancing Physics learning with ChatGPT, bing Chat, and bard as agents-to-think-with: a comparative case study, *SSRN Electron. J.* (2023), <https://doi.org/10.2139/ssrn.4478305>.
- [75] R. Raman, H. Lathabhai, S. Diwakar, P. Nedungadi, Early research trends on ChatGPT: insights from altmetrics and science mapping analysis, *International Journal of Emerging Technologies in Learning (IJET)* 18 (2023) 13–31, <https://doi.org/10.3991/IJET.V18I19.41793>.
- [76] R. Raman, Transparency in research: an analysis of ChatGPT usage acknowledgment by authors across disciplines and geographies, *Account. Res.* (2023), <https://doi.org/10.1080/08989621.2023.2273377>.
- [77] Ahmed, I., Kajol, M., Hasan, U., Datta, P. P., Roy, A., & Reza, M. R. (2023). ChatGPT vs. Bard: A Comparative Study. UMBC Student Collection | Mendeley (n. d.). <https://www.mendeley.com/search/?page=1&query=Ahmed%2C%20I.%2C%20Kajol%2C%20M.%2C%20Hasan%2C%20U.%2C%20Datta%2C%20P.%20P.%2C%20Roy%2C%20A.%2C%20%26%20Reza%2C%20M.%20R.%20%282023%29.%20ChatGPT%20vs.%20Bard%3A%20A%20Comparative%20Study%20UMBC%20Student%20Collection&sortBy=relevance> (accessed February 15, 2024)..
- [78] R. Ali, O.Y. Tang, I.D. Connolly, J.S. Fridley, J.H. Shin, P.L. Zadnik Sullivan, D. Cielo, A.A. Oyelese, C.E. Doberstein, A.E. Telfeian, Z.L. Gokaslan, W.F. Asaad, Performance of ChatGPT, GPT-4, and Google bard on a neurosurgery oral boards preparation question bank, *Neurosurgery* 93 (2023) 1090–1098, <https://doi.org/10.1227/NEU.0000000000002551>.
- [79] P. Humar, M. Asaad, F.B. Bengur, V. Nguyen, ChatGPT is equivalent to first-year plastic surgery residents: evaluation of ChatGPT on the plastic surgery in-service examination, *Aesthetic Surg. J.* 43 (2023), <https://doi.org/10.1093/ASJ/SJAD130>. NP1085–NP1089.
- [80] B. Campello de Souza, A. Serrano de Andrade Neto, A. Roazzi, Are the new AIs smart enough to steal your job? IQ scores for ChatGPT, Microsoft bing, Google bard and Quora Poe, *SSRN Electron. J.* (2023), <https://doi.org/10.2139/SSRN.4412505>.
- [81] E. Felten, R. Seamans, How will Language Modelers like ChatGPT Affect Occupations and Industries? (n.d.). <https://www.cbsnews.com/news/chatgpt-jpmorgan-chase-bars-workers-from-using-ai-tool/> (accessed February 15, 2024)..
- [82] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, M. Glaese, B. Balle, A. Kasirzadeh, Z. Kenton, S. Brown, W. Hawkins, T. Stpleton, C. Biles, A. Birhane, J. Haas, L. Rimell, L.A. Hendricks, W. Isaac, S. Legassick, G. Irving, I. Gabriel, <Iweidinger@deepmind Com>, Ethical and Social Risks of Harm from Language Models, 2021. <https://arxiv.org/abs/2112.04359v1>. (Accessed 16 February 2024).
- [83] D.L. Deadrick, P.A. Gibson, An examination of the research–practice gap in HR: comparing topics of interest to HR academics and HR professionals, *Hum. Resour. Manag. Rev.* 17 (2007) 131–139, <https://doi.org/10.1016/J.HRMR.2007.03.001>.
- [84] Perceived Value of HRM Professional Certification in a Disrupted Marketplace | *Advances in Business Research*, (n.d.). <https://journals.sfu.ca/abr/index.php/abr/article/view/567> (accessed February 16, 2024).
- [85] S. Loufrani-Pedida, B. Aldebert, A multilevel approach to competence management in innovative small and medium-sized enterprises (SMEs): literature review and research agenda, *Employee Relat.* 43 (2021) 507–523, <https://doi.org/10.1108/ER-04-2020-0173/FULL/XML>.
- [86] S. Kis, L. Mosora, Y. Mosora, O. Yatsiuk, G. Malynovska, S. Pobihun, Personnel certification as a necessary condition for enterprise’ staff development, *Manag. Syst. Prod. Eng.* 28 (2020) 121–126, <https://doi.org/10.2478/MSPE-2020-0018>.
- [87] M. Subramony, J.P. Guthrie, J. Dooney, Investing in HR? Human resource function investments and labor productivity in US organizations, *Int. J. Hum. Resour. Manag.* 32 (2021) 307–330, <https://doi.org/10.1080/09585192.2020.1783343>.
- [88] A. Borji, M. Mohammadian, Battle of the Wordsmiths: Comparing ChatGPT, GPT-4, Claude, and Bard, *SSRN Electronic Journal*, 2023, <https://doi.org/10.2139/SSRN.4476855>.
- [89] G.S. Becker, Human capital: a theoretical and empirical analysis with special reference to education, in: *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, third ed., 1994 ii–xxiii, <https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-third-edition>. (Accessed 15 February 2024).
- [90] R.A. Swanson Elwood F Holton III, *Foundations of Human Resource Development*, 2022. [https://books.google.co.in/books?hl=en&lr=&id=XQI\\_EAAAQBAJ&oi=fnd&pg=PP1&dq=Swanson,+R.+A.+\(2022\).+Foundations+of+human+resource+development,+San+Francisco:+Berrett-Koehler+Publishers&ots=T7EhhCRQG\\_&sig=MEB4Z14V5O17Ac82CYdEvDRYmJk](https://books.google.co.in/books?hl=en&lr=&id=XQI_EAAAQBAJ&oi=fnd&pg=PP1&dq=Swanson,+R.+A.+(2022).+Foundations+of+human+resource+development,+San+Francisco:+Berrett-Koehler+Publishers&ots=T7EhhCRQG_&sig=MEB4Z14V5O17Ac82CYdEvDRYmJk). (Accessed 16 February 2024).
- [91] I. Chak, K. Croxson, F. D’Acunto, J. Reuter, A.G. Rossi, J. Shaw, Improving household debt management with robo-advice, *SSRN Electron. J.* (2022), <https://doi.org/10.2139/SSRN.4269908>.
- [92] F. D’Acunto, N. Prabhala, A.G. Rossi, The promises and pitfalls of robo-advising, *Rev. Financ. Stud.* 32 (2019) 1983–2020, <https://doi.org/10.1093/RFS/HHZ014>.
- [93] D. Ulrich, HR OF THE FUTURE: CONCLUSIONS AND OBSERVATIONS, (n.d.).
- [94] K.K. Kapoor, K. Tamilmani, N.P. Rana, P. Patil, Y.K. Dwivedi, S. Nerur, Advances in social media research: past, present and future, *Inf. Syst. Front* (2017) 531–558, <https://doi.org/10.1007/S10796-017-9810-Y>, 2017 20:3 20.
- [95] E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: can language models be too big? *FACCT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021) 610–623, <https://doi.org/10.1145/3442188.3445922>.
- [96] L. Floridi, M. Chiriatti, GPT-3: its nature, scope, limits, and consequences, *Minds Mach.* 30 (2020) 681–694, <https://doi.org/10.1007/S11023-020-09548-1/FIGURES/5>.
- [97] J. Bommarito, M.J. Bommarito, J. Katz, D.M. Katz, GPT as knowledge worker: a zero-shot evaluation of (AI)CPA capabilities, *SSRN Electron. J.* (2023), <https://doi.org/10.2139/ssrn.4322372>.
- [98] A. Srivastava, A. Rastogi, A. Rao, et al., Beyond the Imitation Game: Quantifying and Extrapolating the Capabilities of Language Models, 2022. <http://arxiv.org/abs/2206.04615>. (Accessed 16 February 2024).
- [99] C. Vallance, AI Could Replace Equivalent of 300 Million Jobs-Report, Retrieved April, 27, 2023. - Google Search, (n.d.), BBC News, 2023, [https://www.google.com/search?q=%C2%A0Vallance%2C+C.+%282023%29.+AI+could+replace+equivalent+of+300+million+jobs-report.+BBC+News.+Retrieved+April%2C+27%2C+2023.&client=safari&sca\\_esv=d8934033fde41234&sca\\_upv=1&rls=en&ei=sHOZD.\(Accessed+16+February+2024\).&HFLQL4-EP37CB4AU&ved=0ahUKEwjVvuz84a6EAxW6xTgGHV9YAFwQ4dUDCA8&oq=%C2%A0Vallance%2C+C.+%282023%29.+AI+could+replace+equivalent+of+300+million+jobs-report.+BBC+News.+Retrieved+April%2C+27%2C+2023.&gs\\_l=EGxnd3Mtd2l6LXNlcnAic8KgVmFsbGFuY2UsIEMuICgyMDIzKS4gQUkqY291bGQgcmVwbGFJZSBlcXVpdmFsZW50IG9mIDMwMkCBtaWxsaW9uIGVpYnMtemcmVwb3J0LjBjCQkMgTmV3cy4gUmV0cmldmVkieFwcmLSLCAyNywMjAyMy5IAFAAWABwAHgBKAEAmAEoAeAqEAuAEMyAEA&scient=gws-wiz-serp](https://www.google.com/search?q=%C2%A0Vallance%2C+C.+%282023%29.+AI+could+replace+equivalent+of+300+million+jobs-report.+BBC+News.+Retrieved+April%2C+27%2C+2023.&client=safari&sca_esv=d8934033fde41234&sca_upv=1&rls=en&ei=sHOZD.(Accessed+16+February+2024).&HFLQL4-EP37CB4AU&ved=0ahUKEwjVvuz84a6EAxW6xTgGHV9YAFwQ4dUDCA8&oq=%C2%A0Vallance%2C+C.+%282023%29.+AI+could+replace+equivalent+of+300+million+jobs-report.+BBC+News.+Retrieved+April%2C+27%2C+2023.&gs_l=EGxnd3Mtd2l6LXNlcnAic8KgVmFsbGFuY2UsIEMuICgyMDIzKS4gQUkqY291bGQgcmVwbGFJZSBlcXVpdmFsZW50IG9mIDMwMkCBtaWxsaW9uIGVpYnMtemcmVwb3J0LjBjCQkMgTmV3cy4gUmV0cmldmVkieFwcmLSLCAyNywMjAyMy5IAFAAWABwAHgBKAEAmAEoAeAqEAuAEMyAEA&scient=gws-wiz-serp)
- [100] PwC Global Workforce Hopes and Fears Survey 2022, (n.d.).
- [101] AI Anxiety: the Workers Who Fear Losing Their Jobs to Artificial Intelligence - BBC Worklife, (n.d.) <https://www.bbc.com/worklife/article/20230418-ai-anxiety-artificial-intelligence-replace-jobs> (accessed February 16, 2024).
- [102] P. Niszczoła, S. Abbas, GPT has become financially literate: insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice, *Finance Res. Lett.* 58 (2023) 104333, <https://doi.org/10.1016/J.FRL.2023.104333>.
- [103] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C.L. Wainwright, P. Mishkin, C.Zhang Sandhini Agarwal Katarina Slama Alex Ray John Schulman Jacob Hilton Fraser Kelton Luke Miller Maddie Simens Amanda Askell, P. Welinder Paul Christiano, J. Leike, R. Lowe, Training Language Models to Follow Instructions with Human Feedback, (n.d.).

- [104] D.R. Lewis, The perils of overconfidence: why many consumers fail to seek advice when they really should, *J. Financ. Serv. Market.* 23 (2018) 104–111, <https://doi.org/10.1057/S41264-018-0048-7>.
- [105] M. Broussard, *Artificial Unintelligence: How Computers Misunderstand the World*, Artificial Unintelligence, 2018, <https://doi.org/10.7551/MITPRESS/11022.001.0001>.
- [106] I. Kirkpatrick, K. Hoque, Human resource professionals and the adoption and effectiveness of high-performance work practices, *Hum. Resour. Manag. J.* 32 (2022) 261–282, <https://doi.org/10.1111/1748-8583.12403>.
- [107] G. Latham, *Work Motivation: History, Theory, Research, and Practice 1950-1975: the Emergence of Theory, Work Motivation: History, Theory, Research, and Practice*, 2012, pp. 29–60, <https://doi.org/10.4135/9781506335520.n3>. (Accessed 16 February 2024).
- [108] M. Javidan, P.W. Dorfman, M.S. De Luque, R.J. House, In the eye of the beholder: cross cultural lessons in leadership from project GLOBE, *Acad. Manag. Perspect.* 20 (2006) 67–90, <https://doi.org/10.5465/AMP.2006.19873410>.
- [109] S. Noy, W. Zhang, Experimental evidence on the productivity effects of generative artificial intelligence, *SSRN Electron. J.* (2023), <https://doi.org/10.2139/SSRN.4375283>.
- [110] Akanksha Saxena, The growing role of artificial intelligence in human resource, *EPRA International Journal of Multidisciplinary Research (IJMR)* (2020) 152–158, <https://doi.org/10.36713/EPRA4924>.
- [111] M.F. Gonzalez, J.F. Capman, F.L. Oswald, E.R. Theys, D.L. Tomczak, Personnel Assessment and Personnel Assessment and Decisions Decisions “Where’s the I-O?” Artificial Intelligence and Machine Learning in “Where’s the I-O?” Artificial Intelligence and Machine Learning in Talent Management Systems Talent Management Systems (n.d.). <https://doi.org/10.25035/pad.2019.03.005..>